



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 111500

TO: Michael Borin
Location: CM1/12A01-12D01
Art Unit: 1631
Thursday, January 08, 2004

Case Serial Number: 09/581286

From: Toby Port
Location: Biotech-Chem Library
CM1-6A04
Phone: 308-3534

toby.port@uspto.gov

Search Notes

Dear Examiner Borin,

Here are the results of your search.
Please feel free to contact me if you have any questions.

Toby Port

THIS PAGE BLANK (USPTO)



STIC SEARCH RESULTS FEEDBACK FORM

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher* or contact:

Mary Hale, Information Branch Supervisor
308-4258, CM1-1E01

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library CM1 - Circ. Desk



THIS PAGE BLANK (USPTO)

111500
O'Bryen, Barbara

From: STIC-Biotech/ChemLib
Sent: Tuesday, January 06, 2004 12:29 PM
To: O'Bryen, Barbara
Subject: FW: Search request: 09/581286

-----Original Message-----

From: Borin, Michael
Sent: Tuesday, January 06, 2004 12:23 PM
To: STIC-Biotech/ChemLib
Subject: Search request: 09/581286

Examiner: M.Borin
CM1 12A01
AU: 1631; Mailbox 12D01

Tel.: 305-4506

RE: 09/581286; gingivalis peptides

Please conduct against the commercial and interference protein databases of :

1. polypeptide SEQ ID 424
2. oligopeptide search for a fragment of at least 40 residues of polypeptide SEQ ID 424 .

Thank you

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2004, 17:44:33 ; Search time 61 Seconds
(without alignments)
2646.310 Million cell updates/sec

Title: US-09-581-286A-424
Perfect score: 5298
Sequence: 1 MKRWTLFLCLLTLSIGWAMA.....VGKNQVPSKQVAGIQLSF 1017

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Maximum Match 100%
Listing first 45 summaries

Post-processing: Minimum Match 0%

Database : A_Geneseq_19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5298	100.0	1017	20 AAY34477	Porphorymonas ging
2	5298	100.0	1046	20 AAY34353	Porphorymonas ging
3	5278	99.6	1014	20 AAY34478	Porphorymonas ging
4	232.5	4.4	757	20 AAY34472	Porphorymonas ging
5	232.5	4.4	763	20 AAY34348	Porphorymonas ging
6	227.5	4.3	867	20 AAY34508	Porphorymonas ging
7	227.5	4.3	875	20 AAY34381	Porphorymonas ging
8	215.5	4.1	834	20 AAY34542	Porphorymonas ging
9	215.5	4.1	907	20 AAY34408	Porphorymonas ging

10	202	3.8	926	20 AAY34557	Porphorymonas ging
11	202	3.8	945	20 AAY34421	Porphorymonas ging
12	201	3.8	2057	21 AAB10667	L. mesenteroides a
13	197	3.7	1536	15 AAR63505	Haemophilus high m
14	197	3.7	1536	21 AAB01846	Haemophilus influe
15	194.5	3.7	833	20 AAY34494	Porphorymonas ging
16	194.5	3.7	876	20 AAY34368	Porphorymonas ging
17	192.5	3.6	682	22 ABB52836	Escherichia coli p
18	192	3.6	1536	18 AAW30293	Non-typeable Haemo
19	191	3.6	1536	14 AAR41723	High molecular wei
20	189	3.6	708	20 AAY34501	Porphorymonas ging
21	189	3.6	772	20 AAY34374	Porphorymonas ging
22	186.5	3.5	827	20 AAY34499	Porphorymonas ging
23	186.5	3.5	828	20 AAY34372	Porphorymonas ging
24	183.5	3.5	2835	23 ABB98574	Dextran saccharase
25	181	3.4	1536	14 AAR41725	High molecular wei
26	180.5	3.4	3333	24 ABR39482	L. cuprina pre-lip
27	178.5	3.4	1161	21 AAY52383	Pseudomonas aerugi
28	173.5	3.3	1601	18 AAW30292	Non-typeable Haemo
29	172.5	3.3	1529	14 AAR41732	High molecular wei
30	172.5	3.3	2123	22 AAE00701	Moraxella catarrha
31	171	3.2	2122	24 ABU08784	Moraxella catarrha
32	168	3.2	767	19 AAW98606	H. pylori GHPO 147
33	168	3.2	1338	14 AAR41731	High molecular wei
34	168	3.2	1598	18 AAW30291	Non-typeable Haemo
35	168	3.2	1648	23 ABB54925	Lactococcus lactis
36	167.5	3.2	1386	24 ABB82573	H. influenzae BASB
37	167.5	3.2	1449	24 ABB82570	H. influenzae BASB
38	167.5	3.2	2834	23 ABB97576	Haemagglutinin-lik
39	167	3.2	1974	19 AAW98391	H. pylori GHPO 57
40	166	3.1	848	20 AAY34512	Porphorymonas ging
41	166	3.1	878	20 AAY34385	Porphorymonas ging
42	165	3.1	719	17 AAW01462	NTHI HxuC protein.
43	164.5	3.1	2732	22 ABB52855	Escherichia coli p
44	162.5	3.1	1346	17 AAR91245	VIP2A(a) and VIP1A
45	162.5	3.1	1346	18 AAW19513	B. cereus VIP1A(a)

ALIGNMENTS

RESULT 1
AAY34477
ID AAY34477 standard; Protein; 1017 AA.

AC AAY34477;

DT 20-MAR-2003 (updated)
DT 25-AUG-1999 (first entry)

DE Porphorymonas gingivalis protein PG2.

KW Porphorymonas gingivalis; PG; periodontal disease; gingivitis;
KW vaccine; antigenic.

OS Porphorymonas gingivalis.

PN WO9929870-A1.

PD 17-JUN-1999.

PF 10-DEC-1998; 98WO-AU01023.

PR 10-DEC-1997; 97AU-0000839.

PR 31-DEC-1997; 97AU-0001182.

PR 30-JAN-1998; 98AU-0001546.

PR 10-MAR-1998; 98AU-0002264.

PR 09-APR-1998; 98AU-0002911.

PR 23-APR-1998; 98AU-0003128.

PR 05-MAY-1998; 98AU-0003338.

PR 22-MAY-1998; 98AU-0003654.

PR 29-JUL-1998; 98AU-0004917.

PR 30-JUL-1998; 98AU-0004963.

dis open

SQ		Sequence		1014 AA;	
Query Match		99.6%; Score 5278; DB 20; Length 1014;			
Best Local Similarity		99.9%; Pred. No. 0;			
Matches 1013; Conservative		1; Mismatches 0; Indels 0; Gaps 0;			
QY	4	MTLFFCLLTISIGWAMAQNRTVKGTVI	SSDNEPLIGANVVVGNNTTIGAA	TDLGNTFL	63
DB	1	MTLFFCLLTISIGWAMAQNRTVKGTVI	SSDNEPLIGANVVVGNNTTIGAA	TDLGNTFL	60
QY	64	SVPANAKMLRVSYSGMTTKEVAIANVMKI	VLDPSKVLQVVLGYGTGOKLSTVSGVA		123
DB	61	SVPANAKMLRVSYSGMTTKEVAIANVMKI	VLDPSKVLQVVLGYGTGOKLSTVSGVA		120
QY	124	KVSEKLAEPVANIIMDALOGVAGQVMTTSGDPTAVASVEIHGTGSLGASSAPLYVD			183
DB	121	KVSEKLAEPVANIIMDALOGVAGQVMTTSGDPTAVASVEIHGTGSLGASSAPLYVD			180
QY	184	GMOTSLDVVATMPNDPESMSVLKASATSIYGARAANGVVFIQTKGKMSERGRITFNA			243
DB	181	GMOTSLDVVATMPNDPESMSVLKASATSIYGARAANGVVFIQTKGKMSERGRITFNA			240
QY	244	SYGISQILNTKPLDNMTGDELLDFQVKAQFNGNNTQVQVKOMILAGABDLYGNVDSLK			303
DB	241	SYGISQILNTKPLDNMTGDELLDFQVKAQFNGNNTQVQVKOMILAGABDLYGNVDSLK			300
QY	304	DEYKTLFPVDFNHADWLKALPKTAPTSOGDISFSGSGSGTSYASIGYFDQEGMAREP			363
DB	301	DEYKTLFPVDFNHADWLKALPKTAPTSOGDISFSGSGSGTSYASIGYFDQEGMAREP			360
QY	364	ANFKRYSGRLNFESRINELWLVKANLSGAIANRRSADYFGKYMGSGTGFGLTWPYRNP			423
DB	361	ADFKRYSGRLNFESRINELWLVKANLSGAIANRRSADYFGKYMGSGTGFGLTWPYRNP			420
QY	424	FDVNGDLADVYMYGATRPMSYFAXMRPFSSSHQANVNGFAQITPIKGLTLKAQAG			483
DB	421	FDVNGDLADVYMYGATRPMSYFAXMRPFSSSHQANVNGFAQITPIKGLTLKAQAG			480
QY	484	VDLTNTTSKRPNNPYDSTPLGERERAYRDVSKFTNTAEYKFSIDEKHLDTALMGH			543
DB	481	VDLTNTTSKRPNNPYDSTPLGERERAYRDVSKFTNTAEYKFSIDEKHLDTALMGH			540
QY	544	EYIEYEGDVTGASKGSPESDKMLLSQKGTGNSLSLPEHRAVEAYVILSFFSRFNGYFDKW			603
DB	541	EYIEYEGDVTGASKGSPESDKMLLSQKGTGNSLSLPEHRAVEAYVILSFFSRFNGYFDKW			600
QY	604	MYIDFSVRNDQSSRFSGNRSAMFYSGMFDIYNKFIQESNMLSDURLKMSYGTTCNSE			663
DB	601	MYIDFSVRNDQSSRFSGNRSAMFYSGMFDIYNKFIQESNMLSDURLKMSYGTTCNSE			660
QY	664	IGNYNHQAALVYNNYTEDAMGLSISTAGNPDLSWEKOSQNFGLAAGAFNNRLSAEVDYF			723
DB	661	IGNYNHQAALVYNNYTEDAMGLSISTAGNPDLSWEKOSQNFGLAAGAFNNRLSAEVDYF			720
QY	724	VRTNDMLIDVPMPIYISGFSQYQVNGSMKNTGVDLSLKGTYIQNDKNVYASAFNNYR			783
DB	721	VRTNDMLIDVPMPIYISGFSQYQVNGSMKNTGVDLSLKGTYIQNDKNVYASAFNNYR			780
QY	784	QEITKLPFLGNKMYLPTNGTITWEIGYPSNFYMAEYAGIDKKTGKQLVYFQVQDADGNKY			843
DB	781	QEITKLPFLGNKMYLPTNGTITWEIGYPSNFYMAEYAGIDKKTGKQLVYFQVQDADGNKY			840
QY	844	TTSOYSADLETRIDKSVTPPTTGGFSLGASWKGSLDADFAIVGKWMNNDRYFTENAG			903
DB	841	TTSOYSADLETRIDKSVTPPTTGGFSLGASWKGSLDADFAIVGKWMNNDRYFTENAG			900
QY	904	GLMQLNKDKMLLANWATEDNKETDVPKLGQSPQFQTHLLENASFLRLKNLKLTVLPNSLF			963
DB	901	GLMQLNKDKMLLANWATEDNKETDVPKLGQSPQFQTHLLENASFLRLKNLKLTVLPNSLF			960
QY	964	AGQNVIGGARVYLMARNLLTVTKYKGFDPBRAGNVGNQKQVPSNQYVAGIQLSF			1017
DB	961	AGQNVIGGARVYLMARNLLTVTKYKGFDPBRAGNVGNQKQVPSNQYVAGIQLSF			1014

RESULT 4		AAY34472	
ID	AAY34472 standard; Protein; 757 AA.		
XX	AAY34472;		
AC	AAY34472;		
XX	20-MAR-2003 (updated)		
DT	25-AUG-1999 (first entry)		
XX	Porphyromonas gingivalis protein PG13.		
DE	Porphyromonas gingivalis; PG; periodontal disease; gingivitis;		
XX	vaccine; antigenic.		
KW	Porphyromonas gingivalis.		
XX	Porphyromonas gingivalis.		
OS	Porphyromonas gingivalis.		
XX	W09929870-A1.		
PN	17-JUN-1999.		
XX	10-DEC-1998; 98WO-AU01023.		
XX	10-DEC-1997; 97AU-0000839.		
PR	31-DEC-1997; 97AU-0001182.		
PR	30-JAN-1998; 98AU-0001546.		
PR	10-MAR-1998; 98AU-0002264.		
PR	09-APR-1998; 98AU-0002911.		
PR	23-APR-1998; 98AU-0003128.		
PR	05-MAY-1998; 98AU-0003338.		
PR	22-MAY-1998; 98AU-0003654.		
PR	29-JUL-1998; 98AU-0004917.		
PR	30-JUL-1998; 98AU-0004963.		
PR	04-AUG-1998; 98AU-0005028.		
XX	(CSLC-) CSL LTD.		
XX	Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;		
PI	Ross BC, Rothe LJ, Webb EA;		
XX	WPI; 1999-385613/32.		
DR	N-PSDB; AAX91690.		
XX	Antigenic Porphyromonas gingivalis peptides for preventing		
PT	gingivitis		
XX	Claim 1; Page 450-451; 588pp; English.		
PS	AAX91536 to AAX91801 encode two hundred and sixty six antigenic		
CC	Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to		
CC	AAX34583. AAX91802 to AAX91989 represent PCR primers used in the		
CC	isolation of the PG polypeptides. The PG polypeptides have antibacterial		
CC	activity with a vaccine mechanism of action. The PG polypeptides can be		
CC	used as vaccines especially against Porphyromonas gingivalis. Probes can		
CC	be used to detect Porphyromonas gingivalis in standard hybridisation		
CC	assays. Porphyromonas gingivalis is involved in periodontal disease		
CC	especially gingivitis.		
CC	(Updated on 20-MAR-2003 to correct PR field.)		
XX	Sequence 757 AA;		
Query Match 4.4%; Score 232.5; DB 20; Length 757;			
Best Local Similarity 20.2%; Pred. No. 3.8e-07;			
Matches 1171; Conservative 124; Mismatches 279; Indels 271; Gaps 41;			
QY	1	MKRWTLF-----FLCLTISIGWAMAQNRTV-KGTVISSDNEPLIGANVVVGNNTTIGAA	54
DB	1	MRTKTIFFAIISFIALLSS--LSAQSKAVLTGVSDAETGEPLAGAR-IEVKHTNIVAG	57
QY	55	TDLGNTFL-SVPANAKMLRVSYSGMTTKEVAIA-----NVMKIVLDPSKVLQVVL	107
DB	58	ADAGGHFEIKNLPAQOHTIICSLGYGOKEEVVAIEAGOTKTITSPALRLRTNNLEVVVT	117

RESULT 4

AAV34472

ID AAY34472 standard; Protein; 757 AA.

XX AAY34472;

AC AC

DT 20-MAR-2003 (updated)

DT 25-AUG-1999 (first entry)

XX Porphorymonas gingivalis protein PG13.

XX Porphorymonas gingivalis; PG; periodontal disease; gingivitis;

KW vaccine; antigenic.

XX Porphorymonas gingivalis.

OS WO9929870-A1.

PN 17-JUN-1999.

PD 10-DEC-1998; 98WO-AU01023.

XX 10-DEC-1997; 97AU-0000839.

XX 31-DEC-1997; 97AU-0001182.

PR 30-JAN-1998; 98AU-0001546.

PR 10-MAR-1998; 98AU-0002264.

PR 09-APR-1998; 98AU-0002911.

PR 23-APR-1998; 98AU-0003128.

PR 05-MAY-1998; 98AU-0003338.

PR 22-MAY-1998; 98AU-0003654.

PR 29-JUL-1998; 98AU-0004917.

PR 30-JUL-1998; 98AU-0004963.

PR 04-AUG-1998; 98AU-0005028.

XX (CSLC-) CSL LTD.

XX Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;

PI Ross BC, Rothel LJ, Webb EA;

XX WPI; 1999-385613/32.

DR N-PSDB; AAX91690.

XX Antigenic Porphorymonas gingivalis peptides for preventing

PT gingivitis

XX Claim 1; Page 450-451; 588pp; English.

XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic

XX Porphorymonas gingivalis (PG) polypeptide sequences given in AAY34318 to

XX AAY34383. AAX91802 to AAX91989 represent PCR primers used in the

XX isolation of the PG polypeptides. The PG polypeptides have antibacterial

XX activity with a vaccine mechanism of action. The PG polypeptides can be

XX used as vaccines especially against Porphorymonas gingivalis. Probes can

XX be used to detect Porphorymonas gingivalis in standard hybridisation

XX assays. Porphorymonas gingivalis is involved in periodontal disease

XX especially gingivitis.

XX (Updated on 20-MAR-2003 to correct PR field.)

XX SQ Sequence 757 AA;

Query Match 4.4%; Score 232.5; DB 20; Length 757;

Best Local Similarity 20.2%; Pred. No. 3.8e-07;

Matches 171; Conservative 124; Mismatches 279; Indels 271; Gaps 41;

QY 1 MKEMTLF-----FLCLTISIGWAMAQNRTV-KGTIVISSDNEPLIGANVVVGNNTTIGAA 54

DB 1 MRYKTFFAIISFIALLSSS--LSAQSKAVLTGSVSDAETGEPLAGAR-IEVKHTNIVAG 57

QY 55 TDLDGNTFL-SVPANAKMLRVSYSGMTTKEVAIA-----NVWKIVLDPPSKVLEQVVL 107

DB 58 ADAGGHFEIKNLPAGQHTTICSLGGYQKQEVVAIEAGQTKTISFALRLRNNLEEVVVT 117

QY 108 GYGTGQKLTSGSVKAVSSEKLAEPKPVANIMDALQGVAGQVMTTSGD--PTAVAS-V 164
 Db 118 GTGTRYRLVD-----APVATEVLTAKDIASF--SAPTSSEALLQGLSPSDFGPNLMGSPM 170
 QY 165 EIHGTGSLGASSAPLYIVDGMQTSLDV-----VATWNPNDPESMSVLKXASATSIYGARA 219
 Db 171 QUNGL-----SSKYLILIDGKRVYDVGQADLSRISPDQIERIELVKGAS--SSLYGSDA 225
 QY 220 ANGUVFIOTKKGKMSERGRITFNASYGISQILNTRKPLDNMTGDELLDPQVKAQFAGNQQ 279
 Db 226 IAGVINVTKK-----NTNLSAYTSRISK-----YNDRQ 256
 QY 280 TVQKVKMDILAGAEDLYGNYSLSKDEYKTLFPVDFNHDADWLKALFKTAPTQSQDIFS 339
 Db 257 T-----NTSLDINIG 266
 QY 340 GSGQSTSYASIGYFDQEGMAREPANFKYSGRLNPFESRINEWLKVGANLSAIAANRRA 399
 Db 267 KFSSTNTYF-----FYHTDQWQNSPFEIKKKKG--SGEPVLEETK----- 305
 QY 400 DYFGKYMGSGTFGLVTPRYX--NPPDVNGDLADVYVNYMGATRPMSMTEPYFAKMRPSS 457
 Db 306 ----KTFRAGENQGVQSLSYATNLSFG--NVQY----NKQIFPTPSEKKAYDM 354
 QY 458 ESHQANVN--GFAQITPIKGL--TLKAQAGVD-----ITWTRTSSKRMPPNPDYDSTPLGER 509
 Db 355 DYRALTASLGNTYLPF--NGLHTLSFDVYDRFRFGYLYHDKDSSESLINNQ-----GQT 407
 QY 510 REAY-----RDVSKFTNTAYK--FSIDEKHDLTALMCHYIIEYEGDVICASSKGPES 562
 Db 408 EQPTFPFGQLRKNQDIQRYTAEARGVFTLTPYAQKLTG--GLSYFREE----- 452
 QY 563 DKMLLSQKGTGNSLSLPHRAVEAYLSFFSRFNFGDKWYIDFSVENDQSSRFGSNN 622
 Db 453 -----LISPNLITDRAASTLSA--YVDEWKLDP----- 481
 QY 623 RSANFYSGGMFDIYNKFIQESNWLSDLRKMSYGTGTGNSSEIGNYNHQA----- 671
 Db 482 ---WFNMTAGFRLVHQ-----EFGTRMTPKVSILAKYGPLNFRATYANGYKTPTL 529
 QY 672 --LVTVNNYTEDAMGLSISTAGNPDLSWEKQSFNGL-----AAGAFNNLSAEV 720
 Db 530 KLFARNELT--TWGSHNLYLGNADLKQMSDYIALGLEYNQGPISFSATVYDNELRNLI 587
 QY 721 DFYVVRTNDMLIDVMP-----YISGF--PSQYQNVGSMKNTGYDLSLKGTYIQNKDWNVYA 775
 Db 588 SF-----WDIPTSPHEAAGIKTKQYANIGKARSGLDVLCDASI-----GWGKIL 634
 QY 776 SANFN 780
 Db 635 GAGYS 639
 RESULT 5
 ID AAY34348
 AC AAY34348 standard; Protein; 763 AA.
 XX AAY34348;
 XX
 DT 20-MAR-2003 (updated)
 DT 25-AUG-1999 (first entry)
 XX
 DE Porphyromonas gingivalis protein PG13.
 XX
 KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
 KW vaccine; antigenic.
 OS Porphyromonas gingivalis.
 XX
 FN W0929870-Al.
 XX
 PD 17-JUN-1999.

XX 10-DEC-1998; 98WO-AU01023.
 XX 10-DEC-1997; 97AU-0000839.
 PR 31-DEC-1997; 97AU-0001182.
 PR 30-JAN-1998; 98AU-0001546.
 PR 10-MAR-1998; 98AU-0002264.
 PR 09-APR-1998; 98AU-0002911.
 PR 23-APR-1998; 98AU-0003128.
 PR 05-MAY-1998; 98AU-0003338.
 PR 22-MAY-1998; 98AU-0003654.
 PR 29-JUL-1998; 98AU-0004917.
 PR 30-JUL-1998; 98AU-0004963.
 PR 04-AUG-1998; 98AU-0005028.
 XX
 PA (CSLC-) CSL LTD.
 XX
 PI Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
 PI Ross BC, Rothel LJ, Webb EA;
 XX
 DR WPI; 1999-385613/32.
 XX N-PSDB; AAX91566.
 PT Antigenic Porphyromonas gingivalis peptides for preventing
 PT gingivitis
 XX
 PS Claim 1; Page 308-309; 588pp; English.
 XX
 CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
 CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAX94318 to
 CC AAX94583. AAX91802 to AAX91989 represent PCR primers used in the
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The PG polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis.
 CC (Updated on 20-MAR-2003 to correct PR field.)
 XX
 SQ Sequence 763 AA;
 Query Match 4.4%; Score 232.5; DB 20; Length 763;
 Best Local Similarity 20.2%; Pred. No. 3.8e-07;
 Matches 171; Conservative 124; Mismatches 279; Indels 271; Gaps 41;
 QY 1 MKRMTLF-----FLCLTSSIGWMAQNTV-KGTVISSEDEPLIGANVVVVNTTIGAA 54
 Db 7 MKRTTIFPAIISFIALSSS--LSAQKAVITGSVSDAETGEPLAGAR-IEVKHTNIVAG 63
 QY 55 TDLDCNFTL-SVPANAKMLRVSYSGMTTKVAIA-----NVMKIVLPDSDSKVLQVWVL 107
 Db 64 ADAGGHFEIKNLPAGQHTIICSLGGYQKKEVVAIEAGQTKTISFALRLRTNNLEVVVT 123
 QY 108 GYGTGQKLTSGSVKAVSSEKLAEPKPVANIMDALQGVAGQVMTTSGD--PTAVAS-V 164
 Db 124 GTGTRYRLVD-----APVATEVLTAKDIASF--SAPTSSEALLQGLSPSDFGPNLMGSPM 176
 QY 165 EIHGTGSLGASSAPLYIVDGMQTSLDV-----VATWNPNDPESMSVLKXASATSIYGARA 219
 Db 177 QUNGL-----SSKYLILIDGKRVYDVGQADLSRISPDQIERIELVKGAS--SSLYGSDA 231
 QY 220 ANGUVFIOTKKGKMSERGRITFNASYGISQILNTRKPLDNMTGDELLDPQVKAQFAGNQQ 279
 Db 232 IAGVINVTKK-----NTNLSAYTSRISK-----YNDRQ 262
 QY 280 TVQKVKMDILAGAEDLYGNYSLSKDEYKTLFPVDFNHDADWLKALFKTAPTQSQDIFS 339
 Db 263 T-----NTSLDINIG 272
 QY 340 GSGQSTSYASIGYFDQEGMAREPANFKYSGRLNPFESRINEWLKVGANLSAIAANRRA 399
 Db 273 KFSSTNTYF-----FYHTDQWQNSPFEIKKKKG--SGEPVLEETK----- 311

Db 575 -SPWKSFLPWS---GKAGLSYFAQGNVFNANGGFFTRAPLFGNIYAAGAIIPNDKANME 630
 QY 699 K--QSQNFGLAAGAFNNRLSAEVDYFVVRTNDMLIDVMPYISGFSQYQNVGSMKNTG 756
 Db 631 KVLGTGVGVG---FTNHKNFEFN-----INGYTKWMDRVTSKRIG 668
 QY 757 VDLSLKGTIYQN-----KQNVYASANFNYNRQBITKLPFLGNKYMPLNTGTTWIEIGYPS 812
 Db 669 NEY-----VYLNGVDVHVGCGVEAEVSYPRIQIDLRGMEFLGDW-----TWQ-----NN 712
 QY 813 FYMAEYAGDKTKGKOLWYVPGQVDAGNKNVTTQSADLETRIDKSVTPTTGGFSLGA 872
 Db 713 VSYTSDEAGNETGQDITIKGLHVGDAQAQMTAA--VSADIEL----- 753
 QY 873 SWKGLSLDADFAIVYIGKWMINNDRY--FTENAGGLQLNKKD--KMLINAWTENDKEDTDP- 928
 Db 754 -FKGF-----HVIGKYNFLGKNYAGFNPAENAAQYEADEKEIVESW---KLDPDVL 801
 QY 929 -KLQSQPQPDTHLLENASFLRLKNL--KLTVYVLPNSLFAQONVIG 970
 Db 802 FDLASYNFKGLSLSTTFYFNDNADKRYVSD----ADDNIIG 841

RESULT 7

AA34381
 ID AAY34381 standard; Protein; 875 AA.
 AC AAY34381;
 XX
 XX
 DT 20-MAR-2003 (updated)
 DT 25-AUG-1999 (first entry)
 XX
 XX
 DE Porphorymonas gingivalis protein PG47.
 XX
 KW Porphorymonas gingivalis; PG; periodontal disease; gingivitis;
 KW vaccine; antigenic.
 XX
 OS Porphorymonas gingivalis.
 XX
 PN WO929870-Al.
 XX
 XX 17-JUN-1999.
 XX
 XX 10-DEC-1998; 98WO-AU01023.
 PF
 PF 10-DEC-1997; 97AU-0000839.
 PR 31-DEC-1997; 97AU-0001182.
 PR 30-JAN-1998; 98AU-0001546.
 PR 10-MAR-1998; 98AU-0002264.
 PR 09-APR-1998; 98AU-0002911.
 PR 23-APR-1998; 98AU-0003128.
 PR 05-MAY-1998; 98AU-0003338.
 PR 22-MAY-1998; 98AU-0003654.
 PR 29-JUL-1998; 98AU-0004917.
 PR 30-JUL-1998; 98AU-0004963.
 PR 04-AUG-1998; 98AU-0005028.
 XX
 PA (CSLC-) CSL LTD.
 XX
 PI Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
 PI Ross BC, Rothel LJ, Webb EA;
 PI
 XX
 DR WPI; 1999-385613/32.
 DR N-PSDB; AAX91599.
 XX
 XX Antigenic Porphorymonas gingivalis peptides for preventing
 PT gingivitis
 PT
 XX
 PS Claim 1; Page 346-348; 588pp; English.
 PS
 XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
 CC Porphorymonas gingivalis (PG) polypeptide sequences given in AAY34318 to

CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The PG polypeptides can be
 CC used as vaccines especially against Porphorymonas gingivalis. Probes can
 CC be used to detect Porphorymonas gingivalis in standard hybridisation
 CC assays. Porphorymonas gingivalis is involved in periodontal disease
 CC especially gingivitis.
 CC (Updated on 20-MAR-2003 to correct PR field.)
 XX
 SQ Sequence 875 AA;
 Query Match 4.3%; Score 227.5; DB 20; Length 875;
 Best Local Similarity 19.5%; Pred. No. 1e-06;
 Matches 207; Conservative 152; Mismatches 380; Indels 325; Gaps 55;
 QY 4 MTLFELCLLTSLGMAAQNRT--VKGTVISEDNEPLIGANVVV--VGNTTIGAATDLGDNF 61
 Db 14 LQLFTALLLALGSLAIAQTAVTGVKVIDSEBELIGVSVSTGQASURGVTTMDGGF 73
 QY 62 TILSVPANAKMLRVSYSGMTTKEVAIANYMK-----IVLDPDSKVLQVNVVLGY 109
 Db 74 RFEVPAK-----SVLTFRCVGYATVTRSIGRGSQEDLGTLLDPOAIGLDEIQVI-- 123
 QY 110 GTGQKLTSTVSGSVAKVSEKLAEPVANIIMDALQCVAGMQVMTSGDPTAVASVEIHGT 169
 Db 124 -----ASVVPKDRMTPEVPSNI-----RVADIQA----- 147
 QY 170 GSLGASSAPLYIVDCMQTSLDVVATNPNDPESMSVLKXASATSIYGARAANGVVPQTK 229
 Db 148 -----ASLNV-----EPPELVK--STPSTYTTKSGSGF-----G 174
 QY 230 KKMSEGRITFNASYSQILNTKPLDNMTG-----DELDFQVKAGF----- 274
 Db 175 DGTNVRGFDIVN--FGV--LINGVPVNGMEDGKYVWNSWGLMQASTIQIQRGLGASK 230
 QY 275 -----WGNNTQVQVKDMILAGEDL--YGNVDSLKDEYGTLPFPVFNHDDADWLKALPK 327
 Db 231 LGISSVGGTMMIITKTDTANTGGSAYVGMGN-----DGLHKSFSISGTGMDGWA----- 280
 QY 328 TAPTSQGDISFGSGSGQTSYYASIGYFQEGMARPEANPKRYSGRLNESRINEMWLKVA 387
 Db 281 -----ITTAG-----SHMTGLGYV--KGL-----KGRAFSYFFNVSKKFN--RHTL 318
 QY 388 NLSGAIA-----NRRSADY-----FGKY--YMGSTFGVLTWPRYPNPFVNGDLADVVYMG 438
 Db 319 SLTGFGAPQWENQORSKYSVADYDKYGRHNSQSGYL-----RGELTPAYAYN 367
 QY 439 A--TRPSMTEPYFAKM-----RPFSSSHQANVNGFAQITPIKGLT 477
 Db 368 TYHKPQPSLNHFWMKMDENTSLYTAXYASLATGGRRAYGKNSKWVLIN--YNTGQPVQTK 426
 QY 478 LKAQAGVDITWTRTSSKMPNPNYDSTPLGERRERAY-----RDVSKSFNTAEYK-- 528
 Db 427 VTPDGLIDYDAVLAANAASNGSEAFALGNSHKNWFGLLSFFKKKKLSSLTAGYDGR 486
 QY 529 -FSIDKHDLTALMGHE--YIEYEGDIVGASSKGFESDKMLLSQGTGNSLSLPEHRVAE 586
 Db 487 YVRGDHYDKITDGLGSSYYIEDPKTKLAYHAG-----QQLKVGDIVN--RDYTCG 535
 QY 587 YAYLSFFSRFNYGFDKWM--YIDFSVRND--QSSRFGSNNRSAMFYSGVGMFDIYNKFIQ 642
 Db 536 IMWHGLPAQMEHS--SEWIDAFVSGSINLYELRNHNHNGSKSTGYLPGV----- 582
 QY 643 ESNWLSDLRLKMSYGTTCGNSGIGNVNHQALVTNNYTEDAMGLSISTAG-----NPDLSWE 698
 Db 593 -SPWKSFLPWS---GKAGLSYFAQGNVFNANGGFFTRAPLFGNIYAAGAIIPNDKANME 638
 QY 699 K--QSQNFGLAAGAFNNRLSAEVDYFVVRTNDMLIDVMPYISGFSQYQNVGSMKNTG 756
 Db 639 KVLGTGVGVG---FTNHKNFEFN-----INGYTKWMDRVTSKRIG 676
 QY 757 VDLSLKGTIYQN-----KQNVYASANFNYNRQBITKLPFLGNKYMPLNTGTTWIEIGYPS 812

Db 677 NEY-----VYLNGVDVHCGVEAEVSYPRIQIRDLRGMSLGDW-----TWQ-----NN 720

Qy 813 FYMAEYAGIDKTKGKQLWYVPGQVDAGNKVTTTSQYSADLETRIDKSVTPITGGPSLGA 872

Db 721 VSYTSYDEAGNETGQDITYIKGLHVGDAQAQTAA-VSADIEL----- 761

Qy 873 SWKGLSLDADPAYIVGKWMINNDRY--FTENAGGLMQLNKD-KMLLNAWTEDNKETDVP- 928

Db 762 -FKGF-----HWIKYFNFLGKYAGFNPATRNAQQYEADGKEIVESW-----KLPDVL 809

Qy 929 -KLGQSPQFDTHLENASFLRLKML-KLTVYLPNSLFAGQNVIG 970

Db 810 FDLASYNFKLGLSLTTFYFNMDNVADKRYVSD-----ADDNIIG 849

RESULT 8

AAV34542

ID AAV34542 standard; Protein; 834 AA.

AC AAV34542;

XX 20-MAR-2003 (updated)

DT 25-AUG-1999 (first entry)

XX Porphyromonas gingivalis protein PG71.

XX Porphyromonas gingivalis; PG; periodontal disease; gingivitis;

KW vaccine; antigenic.

XX Porphyromonas gingivalis.

OS WO9929870-A1.

PN 17-JUN-1999.

XX 10-DEC-1998; 98WO-AU01023.

XX 10-DEC-1997; 97AU-0000839.

PR 31-DEC-1997; 97AU-0001182.

PR 30-JAN-1998; 98AU-0001546.

PR 10-MAR-1998; 98AU-0002264.

PR 09-APR-1998; 98AU-0002911.

PR 23-APR-1998; 98AU-0003128.

PR 05-MAY-1998; 98AU-0003338.

PR 22-MAY-1998; 98AU-0003654.

PR 29-JUL-1998; 98AU-0004917.

PR 30-JUL-1998; 98AU-0004963.

PR 04-AUG-1998; 98AU-0005028.

XX (CSLC-) CSL LTD.

XX Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;

PI Ross BC, Rothel LJ, Webb EA;

XX WPI; 1999-385613/32.

DR N-PSDB; AAX91760.

XX Antigenic Porphyromonas gingivalis peptides for preventing

PT gingivitis

XX Claim 1; Page 535-536; 588pp; English.

XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic

CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAX34318 to

CC AAX34583. AAX91802 to AAX91989 represent PCR primers used in the

CC isolation of the PG polypeptides. The PG polypeptides have antibacterial

CC activity with a vaccine mechanism of action. The PG polypeptides can be

CC used as vaccines especially against Porphyromonas gingivalis. Probes can

CC be used to detect Porphyromonas gingivalis in standard hybridisation

CC assays. Porphyromonas gingivalis is involved in periodontal disease

CC especially gingivitis.

CC (Updated on 20-MAR-2003 to correct PR field.)

XX

SQ Sequence 834 AA;

Query Match 4.1%; Score 215.5; DB 20; Length 834;

Best Local Similarity 20.1%; Pred. No. 6.3e-06;

Matches 199; Conservative 144; Mismatches 337; Indels 309; Gaps 48;

Qy 1 MKEMTLFFLCILTSIGWMAQNR-----VKGTVISSDNEPLIGANV-----VVVGNIT--I 51

Db 1 MKRIVLSSFLFVLISLSLAQNNLTLDVHISGITIKDASSGEPVYATVIRLTGADITQVF 60

Qy 52 GAATDLGNTLSVPANAKM-LRVSYSGMTTKEVAIA-----NMKIVLDPDSKVLQEV 104

Db 61 RQVTDGNGYFVIGLPAAPSVHLTASFVGMKTHTMQISRGNGQHDIKSIDISLE----- 115

Qy 105 VVLGYGTGQKLSVSGSVA-----KVSSEKLA-----EKPA-----NINDALQGGVAGQVMT 153

Db 116 -----DKQLSTVTVSAARPLVKMEIDRLSYNMKDDPAAKTNLLEMLRN----- 159

Qy 154 TSGDPTAVASVEIHGTGSL-----GASSAPLYIVDGMOTSL-----DVAATMNPDESMS 204

Db 160 -----VPLVTVDGQGNIOVKSSNFKHL-NGRSTWVSSNPKEVFRSIPAHITIKRVE 211

Qy 205 VLKASATSIYGARAAANGVVFIOTKGKMSE--RGRITFNASVIGISQILNTKPLDNMTG 262

Db 212 VITDPGVK--YDAEGTSAILDIVTERGKKLEGYSGSITASVS----- 251

Qy 263 DELLDFQVAGFWGNQTVQVKMDILACAGEDLYGNYDSLKDEYKTLFPVDENHDADWL 322

Db 252 -----NNPTAN-----GSI-FLTAKSGKVGLTNTNY----- 277

Qy 323 KALFKTAPTSQGDISFSGSGSQTSY-----ASIGYFDQEGMAREPANFKRYSRLNPESR 378

Db 278 -----YGGKKSRYFERTTSMLOTEBKGQETEGGHFNALLSFEID 322

Qy 379 INEWLKVGAN--LSGAIANRRSADYFGKYMGGSTFGV-----LTMPRYNNPFDVNGDLAD 432

Db 323 SLNLFVGGNVLWEMTTDRNSVE--KSFAGSNLMSYIDRKLKTQMDAGSYELNAD-- 376

Qy 433 VYMYCATRPS--WTEPYFAKMRPFSSS-----HOANVNGPAQITPIKGLTLKA 480

Db 377 --YQHSTRLPGLLELLTVSYRFTNPNNSSETFIDQWKDPLNTANTIOYA-----GOHSKS 428

Qy 481 QAGVDITNTRTSSKMPNNPYDSTPLGERR-----ERAYRDVSKSFTNTAEYKFSIDEK 534

Db 429 DAGNDEHTAQVDYTR-----PLGQAHSLEAGLKYIYRHA-----TSDPLYIRPS 473

Qy 535 HDLTALMGHEYIEYEGD-----VIGASSKGF-----ESDKMLMLSQSK 572

Db 474 EDAPWQPGSLYAQNPSNGKFRHDQYIGAAYAGYNYRQDQVSLQTGLURVSSRLKAL----- 529

Qy 573 TGNLSLPEHRVAEYAYLSF--FSRNYGPD-----KMYIDFSVR----- 611

Db 530 -----FPENAAADFSHNSFDWVPQLTLGYTTPSPMKQLKLVNFRIORPAIGOLNRYRLQ 593

Qy 612 -NDOSRFRGNN--RSAMFYSVGGMFDIY-----NKFIQESNMLSRLKLM 654

Db 584 TNDYQVQYGNPDLKSEKRRHHVGLSVNYQYAKWMLTASLDYDFCNNAIONYTFDSDPANPL 643

Qy 655 SVGTTCNSEIGNYNHOALVTVANNYTEDAMGLSISTAGNPDLSWEKOSQFNFLGAAGAPNN 714

Db 644 FHQTYGN--IGREHSPSLNTYAMYT--PAVWVRIMLNGNIDRTFKSE-----ALGIDVN 694

Qy 715 RLSEAEDFVYRTNMDLIDVPMPIYISGFPSQYQNVGSMKNTGVDLSLKGTYQN----- 768

Db 695 SMSGMV-----YSLGMFTLPKDWTVNLFGYHGRSYQTKYD-----GNVFNIGIAQK 744

Qy 769 ---KDMNVYASANNFYNNRQETIKLPGLNKYMLPNTGTIWEIGYPNPSFYM--AEYAGIDKK 824

Db 745 LFDKKLURVLSAM-NIHA-----KYSTWKSRTI-----GNGFTIYSENAIQRS 787

Qy 825 TGKQLWYVFGQVDADGNKVTTSQYSADLE 853

Db 788 VLSLSLYSFGKMTQVRKVERTIVNDLAK 816

KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
 KW vaccine; antigenic.
 XX Porphyromonas gingivalis.

OS Porphyromonas gingivalis.

XX WO9929870-A1.

XX 17-JUN-1999.

XX 10-DEC-1998; 98WO-AU01023.

XX 10-DEC-1997; 97AU-0000839.

XX 31-DEC-1997; 97AU-0001182.

XX 30-JAN-1998; 98AU-0001546.

XX 10-MAR-1998; 98AU-0002264.

XX 09-APR-1998; 98AU-0002911.

XX 23-APR-1998; 98AU-0003128.

XX 05-MAY-1998; 98AU-0003338.

XX 22-MAY-1998; 98AU-0003654.

XX 29-JUL-1998; 98AU-0004917.

XX 30-JUL-1998; 98AU-0004963.

XX 04-AUG-1998; 98AU-0005028.

XX (CSLC-) CSL LTD.

XX Agilus CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;

XX Ross BC, Rothel LJ, Webb EA;

XX WPI; 1999-385613/32.

XX N-PSDB; AAX91775.

XX Antigenic Porphyromonas gingivalis peptides for preventing

XX gingivitis

XX Claim 1; Page 550-551; 588pp; English.

XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic

XX Porphyromonas gingivalis (PG) polypeptide sequences given in AAX91536 to

XX AAX91801. AAX91802 to AAX91899 represent PCR primers used in the

XX isolation of the PG polypeptides. The PG polypeptides have antibacterial

XX activity with a vaccine mechanism of action. The PG polypeptides can be

XX used as vaccines especially against Porphyromonas gingivalis. Probes can

XX be used to detect Porphyromonas gingivalis in standard hybridisation

XX assays. Porphyromonas gingivalis is involved in periodontal disease

XX especially gingivitis.

XX (Updated on 20-MAR-2003 to correct PR field.)

XX Sequence 926 AA;

XX Query Match 3.8%; Score 202; DB 20; Length 926;

XX Best Local Similarity 20.5%; Pred. No. 6.1e-05;

XX Matches 221; Conservative 145; Mismatches 349; Indels 364; Gaps 62;

QY 6 LFFCLLTSIGWA--MAQNRVTGKTVISSDNEPLIGANVVVGNT---TIGAAATLDG 59

DB 7 LLLSILLASLGFVLEAQAQAGVAGRVLDDEGN-FMIQANVLQVQSTQVAVAGATNEKG 65

QY 60 NFTLSVPANAK-MLRVSYSGMTT--KEVATANWKIVL-----DPDSKVLQGVVV 106

DB 66 LFSLKTSQEGDYILRVSYGVYTHDEKISLRNGQTITLKOISMNEDARLLQSVTVQA 125

QY 107 -----LVGTG-----OKLS-----TVSG-SVAKV---SSEKL 130

DB 126 EVVRNTLFEAGSYTVAQASIEELIKLPAGEISDGKITEKIDSKILVDGKEFF 185

QY 131 AEKPVANIMDALQGVAGMVMVT-----TSGDPTAVASVEI-----HGTGS 171

DB 186 SKDPQVAIKNL-PADVNVKQVNLKLSLRMSGPDGDEETVNLTVKPEKKGLFGLQ 245

QY 172 LGASSAPLYIVDGMQTSLDVVATNPNDPMSVVLKQASATSIYGARAANGVFIQTKG 231

DB 246 AGYTDQRYMAG-----NVRFDG-----NKQWTLIGSANNNTNMGFSEMD-- 287

QY 232 KMSEGRITF-----NASYGISQILNTKPLDNNMTGDELL 266

DB 288 --SEMGSMTFPQGGRRGFGNSGGTSSMLGNFSEVFPSSALNT-----GGDARY 338

QY 267 DFOVKAGFWGNQTVQVKMILAGAEGLYGNVDSLKDEYKTLFPVDFNHDA-----D 320

DB 339 GYNDKA-----IETTKRVENILAEG-----NTYMD-----NILERSFSHNCQARPMQ 382

QY 321 WLKALPKTAPTSGQDISFS--GSGQTSYYA--SIGYFDOEGWAREPA---NEKRYSGRL 373

DB 383 W-KPSERTEVFVFPDLISKIDGFFNDTYETKDATGISINKGSIHQTQGNNF--RLNGEL 440

QY 374 NFESRIN-EWLKVGANLSGAIANRRSADYFGKYMGSGTF-----GVLTPRYNPFVNV 427

DB 441 DISHKLNDGRTISASVSGLTDED-----GDGIYQAVLQSVETNQKQFN----- 485

QY 428 GDLADVYMYGATRPSTPEP---YFAKM---RPFSESHQANV-----NGFAIITPIKG 475

DB 486 -DNSNLQY---RLRLSYVEPLGKNYFAQAILNRRFSRRNSDREVRLGDDGQYSI----- 536

QY 476 LTLKAQAGVDITNTRTSSKRMNNPYDSTPLCERRERAYRDVSKSFTNTAEYKFSIDEXH 535

DB 537 --LDSQYGLSYNEFTQYRIGLN-----LKKIAKTWDYTV--GFNVDP-- 575

QY 536 DLTALMGHEVIEYEGDVIGASSKGFESDKMLLSQKGTGNSLSLPEHRAEYAYLSFFSR 595

DB 576 -----NRTVSYR-----SVAGVEQDKLAF-----NRV-----NLSPLMR 604

QY 596 FNYGPDKW--MYIDFSVRNDQSRFGSNRRSAWFYSGVMFDIYNK-FIOESNWLSDLRL 652

DB 605 INYKPSRTTNLRVDYGRRTQPS-----INQIAPVQDITNPLFVTEGN-----PGL 650

QY 653 KMSYGT-----GNSE-----IGNYNHQAALVTNNYTTEDAMGLSISAGNPDL 696

DB 651 KPSYNNVMAMFSDFDAKSRAPNIVFGNYTFDDIVP-NTHYDPSGTGIRTTREYASGT 709

QY 697 WE-----KQSQFNFLAAGAFNRLSAEVDYFVYRTTNDMLIDVMPYISGFFSQ 745

DB 710 WQANLHGTLSLPLKNRFAFRM---SLFNRLAEGQSFINDDKNKAL-----SFRTR 757

QY 746 YQNVGSMKNTGVDSLKGTI--YQ-----NKDMVYASANPNYNRQEIITKLFFG--LN 794

DB 758 ERLTLTYRNWIDTSIGNGIFYMANNLSGQKDSRTY---DFGNYQVALTLPYGRID 814

QY 795 KYMLPNTGTIWEIGYENSFMAEYAGIDKKTGQLWVPCQVDADGNKVTTSQYSADLET 854

DB 815 SDVEYNTNS---GYSGGFLDEW-----LW-----NASLSY 842

QY 855 RI--DKSVTPPTIG-----GFSLGASWKGLSLDADPAVIVGKWMINNDRY-FTENAGG 904

DB 843 SFLRDKAGTLRVNGYDILQQRSSISRSASAINIEESMSNTIGRYVMVDFIYRFNATSGG 901

RESULT 11

AAV34421

ID AAV34421 standard; Protein; 945 AA.

XX AAV34421;

XX 20-MAR-2003 (updated)

DT 25-AUG-1999 (first entry)

XX Porphyromonas gingivalis protein PG83.

XX Porphyromonas gingivalis; PG; periodontal disease; gingivitis;

KW vaccine; antigenic.

XX Porphyromonas gingivalis.

XX WO9929870-A1.

XX 17-JUN-1999.

XX

DR N-PSDB; AAA97904.
 XX New nucleic acid encoding recombinant Leuconostoc mesenteroides
 PT alternan sucrose protein and methods of alternan and fructose
 production
 XX
 XX Claim 1a; Page 30-36; 64pp; German.
 CC This invention describes a novel nucleic acid molecule (I) encoding an
 CC alternan sucrose (E.C. 2.1.1.140 - an enzyme, that belongs to the
 CC glucosyltransferase group). The recombinant, purified alternan sucrose
 CC gene is useful for the fermentative production of alternan (a
 CC carbohydrate) and/or fructose by secreting the enzyme into a
 CC saccharose-containing culture medium. Alternatively, the enzyme is
 CC contacted with a saccharose-containing solution. The alternan and/or
 CC fructose is then isolated from the medium. Cosmetic products or
 CC foodstuffs containing alternan can be produced. Recombinant production of
 CC alternan sucrose is advantageous as it provides a cost effective means of
 CC producing fructose for high fructose containing syrups, production of
 CC which previously has been achieved by costly production from maize
 CC starch. This sequence represents the Leuconostoc mesenteroides alternan
 CC sucrose protein which is described in the method of the invention.
 XX
 SQ Sequence 2057 AA;
 Query Match 3.8%; Score 201; DB 21; Length 2057;
 Best Local Similarity 19.4%; Pred. No. 0.00023;
 Matches 232; Conservative 159; Mismatches 414; Indels 388; Gaps 59;
 27 GTVSSDNBNFLGANYVVVGNNTTIGAAATLDGNTLSVPANAKMLRVSYSGMTTKEVAI 86
 54 GTNDSGEKKVVPSTNDSLKQGTDFWYDSGN-----RVD----- 90
 87 ANVMKIVLDPDSKVLQVWVLYGYGTGQKLSVTSGSAKVSSEKLAEK- PVAN--IMDALQ 143
 91 -----QKTNQILLTAELQKKNKKNLSVISDDTSKDDENISKQTKIANQQTVDATAK 142
 144 GOVAGMQVMTTS--GDPATAVSEIHG-----TGLSASAPLIYVDMQTSIDVVATMN 196
 143 G-----LITSNLSDPITGTHYENHNGYFVYIDASKQVTLQNDIGNLOYFD----- 189
 197 PNDPESMLKDSATSIY-----GARAANGVVFQTK-----KGRMSERGRITFNASYG 246
 190 DNGYQVKGSPRVNGKHYPDSVTGKASSNVDIVNGKAQGVDAQGNLKKSYVADSSGQT 249
 247 ISQILNTKPLDNMTGDELDF-----QVAGFWG--NQTVQVKVDM--ILGAEDLYG 297
 250 YFDGNGQPLIGLQIDGNLQYFNQGVQIKGGFQDVNNKRIYPAPTGNNAVANTEIING 309
 298 -----NYDSLKDEYKTLFPVDFNHDADWLKALFKTAPTSQGDISFSGSGQTSYAS 350
 310 KLGORDANGQVRKAFPSKVAGNTFFPDANGVML-----TGLQTIS-----GKTY--- 354
 351 IGYDFQEGMAREPANFRYSGRLNFESRINEWLKVGANLSG-----ALANR 396
 355 --YLDEQHLR-----KNYAGTFN-----NQPMYFDATGAGKTAIEYQDFDQGLVSQNE 402
 397 RSADYFGKYNGS---GTFGLVTPRYNPPDV--NGDLADVYMYGATRPSMTEPYFAK 451
 403 NTPHNAKSYDKSFFENVGVLATADTWYRPTDILKNGD-----TWTASTE---TD 449
 452 MRP-----FSSESHQANVNGFAQITPIKGLTLKAQAGVDITNTTSSKRPNN----- 499
 450 MRPLMTWPDQKQANLYNFMSS---SKGL-----GITTYTAATSKTLNDAAAFVIQT 500
 500 PYDSTPLGERRERAYRVDSKSFNTAE--YKFSIDEKHDLTALMGHEIYIEYEGDVIGASS 557
 501 AIEQOISLKSSTEWLRDAIDSFVKTOANNKQTEDEAFDGLWLGQGLAYQDD-----SH 556
 558 KGPESDKMLLSQKQTKNSLSLPHRVAEYAYLSFFSRFNYGPKWYIDFVSRNDOSSR 617
 557 RTENTD-----SGNRRKLRQPI-----NIDGS---KDTTDG 585

QY 618 FGSNRSARWFYSVCGMFDIYNKFIQ--ESNWLSDLRLKMSYGT-TGNSEIGTINHOALVT 674
 DB 586 KGSE-----FLLANDIDNSNPVQAEQLNLHL---MNFGSITGNNDANFEDGIRVDA 636
 QY 675 VNNYTEDAMGLS-----ISTAGNPDLS--WE 698
 DB 637 VDNVDADLLKIAGDYFKALYGTDKSDANANKHLSILEDWNGKDPQYVNVQOQNAQLTMDYT 696
 QY 699 KQSQNFGLAAGAFNNRLS---AEVDFFYR--TTNDMLIDVPMPIYISGPFQYONVGSXK 753
 DB 697 VTSQFGNSLTHGA--NNRSNMWVFLDTGYLYNGDLNKKIVDKNPNSGTTLVNRANSQDTK 755
 QY 754 -----NTGVDLSLKGTYIQNKDMNVYASANFNRYNRQETIKL 789
 DB 756 VFNYSFVRAHDYDAQDPIRKAMIDHGIKNMQDTFTFQ---LAQGMERYKQDQENPSG 812
 QY 790 FFLGINKYMLPNTGTIWEIGYPSFYMAEYAGI--DKTKQKQLWYVPGQVDADGNK----- 842
 DB 813 FKKYDYNLPSA-----YAMLLTNKDTVPRVY--GDMYLEGGQYMEKG 854
 QY 843 -VTTSQYSADLETRIDKSVTPPITGGFSLGASKWGLSL---DADFAYIV--GKWMINNDR 896
 DB 855 TIYNPVISALLKARIKY-----VSGGTWATDSSGKDLKDGTDLLTSVRFGRGIMTSDQ 909
 QY 897 YFTE-----NAG-GL-----MOLNKDKML-----LNAW 918
 DB 910 TTTQDNSQDYKNOGIGVIVGNPNPDLKLNNDKTIILHMGKAHKNQLYALVLSNDSGIDVY 969
 QY 919 TEDNKETDVPKLGQSPQFDTLHLENASFLRLKNLKLTYVLPNSL-----FAGONVIGGA 972
 DB 970 DSDDK--APTLRNTDNGDLIFHKHTNTFVKQDGTIINEMKGSNLALISGLYGVWVPVGA 1026
 QY 973 RYVLMARNLLTVTK-----YKGF-----DPEAGNV 998
 DB 1027 SDSQDARTVATESSSNDGVSFHSNAALDSNVIEGFSNFQAMPTSPQSTNV 1079
 RESULT 13
 AAR63505
 ID AAR63505 standard; Protein; 1536 AA.
 XX AAR63505;
 XX AC AC
 XX 25-MAR-2003 (updated)
 XX 25-JUN-1995 (first entry)
 XX Haemophilus high molecular weight protein HMW1.
 XX High molecular weight protein; HMW1, protective vaccine; otitis;
 XX sinusitis; bronchitis; Hib.
 XX Haemophilus.
 XX WO9421290-A1.
 XX 29-SEP-1994.
 XX 15-MAR-1994; 94WO-US02550.
 XX 16-MAR-1993; 93US-0038682.
 XX (BARE/) BARENKAMP S J.
 XX (SGEM/) ST GEME J W.
 XX Barenkamp SJ, St GEME JW;
 XX WPI; 1994-316665/39.
 XX N-PSDB; AAQ72293.
 XX New immunogenic high mol. wt. proteins of non typeable
 PT Haemophilus - useful in protective vaccines
 XX
 XX Claim 2; Page 31; 127pp; English.

775 STGSLRFTKSTGKSTGFSISKDTLTLNATG-----GNITLLQVEGTDMGKGI-VA 825

833 PQCVADAGNKVTTQSYSDADLETRDKSVT-----PPITGGFSLCASW 874

826 KKNITFEKGNITFGSRKA--VTEIEGNVTINNANVTLLGSDPDNHQKPLTIKKOVIINS 883

875 KGLSLDADPAIVGVKWMIND-----RYFTENAGGLMQLNKKMLNANTENDKETDVP 928

884 GNLTAGGNIVNIAGNLTVESNANFKAITNFTFNVGGLFD-----NKGNSNIISA 932

929 KLGQSPQPDTHLLENASFLRLKNLKLTIVLPNSLFAQGNVIGGARVYLMARNL-----L 962

933 KGG-----ARFQIDNRSKLSITNSSSTYRTIISG-----NITKNKNDL 972

983 TVTKYKGFDP--AGSNVGNKQ 1002

973 NITN-EGSDTEMQIGDVSQKE 993

RESULT 14

AAB01846

ID AAB01846 standard; Protein; 1536 AA.

XX AAB01846;

XX

XX 11-SEP-2000 (first entry)

XX Haemophilus influenzae strain 12 HMW1A protein, SEQ ID NO:67.

DE

XX

XX HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight; non-typeable Haemophilus influenzae; NTHi; non-encapsulated; recombinant production; Escherichia coli; antibacterial; vaccine; human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis; detection; diagnosis.

XX

XX Haemophilus influenzae strain 12.

XX

XX WO200020609-A2.

PN

XX

PD 13-APR-2000.

XX

XX 07-OCT-1999; 99WO-CA00938.

PF

XX

XX 07-OCT-1998; 98US-0167568.

PR

XX

XX 08-DEC-1998; 98US-0206942.

PR

XX

PA (CONN-) CONNAUGHT LAB LTD.

XX

XX Loesmore SM, Yang Y, Klein MH;

PI

XX

XX WPI; 2000-303789/26.

DR

XX

XX N-PSDB; ARA52195.

XX

XX

XX Nucleic acid molecule for producing recombinant high molecular weight proteins of Haemophilus which are used as a vaccine to provide protection against Haemophilus induced diseases in humans -

PT

XX

PS Example 16; Fig 28A-Q; 307pp; English.

XX

XX The invention relates to the recombinant production of Haemophilus influenzae high molecular weight (HMW) proteins in Escherichia coli. The expression construct used to effect recombinant expression comprises a promoter functional in E. coli (e.g., the T7 promoter) operably linked to a modified hmwABC operon from a non-typeable (non-encapsulated) H. influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA, hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins and the hmwB and hmwC genes encode accessory proteins which are responsible for post-translational processing and secretion of the HMWA proteins. The modified hmwABC operon used in the expression construct of the invention contains an A gene modified such that it encodes only the mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198) and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae

CC strains Joyce, K1, K21, LDC2, PMH1, 15 and 12. The nucleic acids and
CC vectors are used for the production of recombinant H. influenzae HMW
CC proteins which can be used as vaccines to mediate a humoral or
CC cell-mediated immune response to provide protection against diseases in
CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC antigens in immunoassays for detecting antibodies against Haemophilus,
CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC HMW proteins can be used to isolate and clone hmw genes from other
CC non-typeable strains of Haemophilus via hybridisation reactions. The
CC present sequence represents an HMW protein from a non-typeable strain of
CC H. influenzae.
XX
SQ Sequence 1536 AA;

Query Match 3.78; Score 197; DB 21; Length 1536;
Best Local Similarity 18.74; Pred. No. 0.00028;
Matches 206; Conservative 151; Mismatches 377; Indels 368; Gaps 51;

QY 25 VKGTVISSENEPLIGANVVVG-----NTTGAATDLCNFTLSVPANAKMLRVSYSGM 79
DB 136 LKG-ILDSNGQVFLNPGITIGKDAINTNGFTASTLD-----ISNENIYARNFTPE 187

QY 80 TTREVAIANVMK---IVLDPD-----SKVLEQVVVLGYGTGQKLSVSGSVAKVSSEK 129
DB 188 QTKDKALAEIVNHGLITVKGXGWNLIGGVKXKNEGVI-----SVNGGSIISLLAGQK 238

QY 130 LAEKPVAN-----IMDALQOVAGMOWTTSQDPTAVASVEIHGTGSLGASSAPLYVDG 184
DB 239 ITISDIINPTITYSIAAPENAEVNLGDFAFKGGNINVRAATIRNOGKLSAD-----289

QY 185 MQTSLDVATMNPDPFESMVLKASATSIVGARA-----NGVVFIOTKKKMSRGRIT 240
DB 290 -----SVSKDKSGNIVLSAKEGEABEGVISAQNOQAK---GGRK- 326

QY 241 FNASYGISQILNTPKLDNMTGDE-----LLDFQVQKAG---FWGNNQTVQVKOMILA 290
DB 327 -----MITGDKVTLTKGDAVIDLSGEGGETVYLGDERGEGKGIQLA 368

QY 291 GAEDLYGNVD---SLKDEYKTLF-----PVDNHDADWLKALFKTAPTQGDISFSGS 342
DB 369 KKTSLKSGSTINVSKEKGRAIVMGDIALIDGNNIAQ-----GSGDIATKGGF 417

QY 343 QGTS-----YVASIGYFQDEGMAREPANFK---RYSGRINPESRINEWLKVGANLSGAIA 395
DB 418 VETSGHDLPIKDAIVDAKWLDPDVSINAEATAGSN-TSEDDVETGSGNSASTPKRN 476

QY 396 RRSADYFG---KYMGSGTGFVLTPRYNPPFVNGDLADVYYMYGATRPSMTPEYPAKM 452
DB 477 KEXTLTNTTLESILKGTFTVNT-----ANQ 503

QY 453 RPFSSSHQANVGPAQITPIKGLTL-----KAQAGVDITWRTSSKMPNPNVDSTPLGE 508
DB 504 RIYVNSINLS-NG-----SLTLWSEGRSGGVEINNDITT-----GD 540

QY 509 RRERA-----YRDVSKFTNTAEYKFSIDKDHDLTALMG-HEYLEYEGDVGASSKG 559
DB 541 DTGANTITVSGGVVDVHKNISLGAQGNINITAKQDIAFEKSGNQVITGGTITSGNQKG 600

QY 560 FESDKMLLSQGTGSLSLPHEHRAVEAYLSLSPFSRNFYDFKWMYIDFVRNDQSSRF 619
DB 601 FRFNNVSL---NGTSGSLQFTTKRTKNYATNKE-----G 633

QY 620 SNRRAWFYVGVGMFDI-----YKPIQESNW-LSDLRLKMSYGTGTGNSBTGNYN 668
DB 634 TLN-----ISGKVNISMVLPKNSGYDKFKRTYWNLTSLNV-----SESGEFN 677

QY 669 HQALVTWNNYTEDAMGLSISTAGNPDLSEKQSFNFGLAAGAFNNRLSAEVDYFVRTTN 728
DB 678 -----LTIDSGSDSAGTLTOPNVLNGLISFNKDTTFNV-----ERNARVNF-----718

QY 729 DMLIDVMP-----YISGFFSQQYQ-NVGSMMKNTGVDLSL-----KGTIYQNKDNV 773
DB 729 DMLIDVMP-----YISGFFSQQYQ-NVGSMMKNTGVDLSL-----KGTIYQNKDNV 773

Db 719 ----DIKAPIGINKYSSLNYSFNGNISVSGGSDVFTLLASSSNVQTPGVINSKYFNV 774
QY 774 YASANFNYNRQETTKLFFGLNKYMLPN-TGTIWEIGYPNSFYMAEYAGIDKTKGKQLWYV 832
DB 775 STGSSLRPKTSKTKGFGIEKDLTLNATG-----GNITLLQVEGTDGMIGKGI-VA 825
QY 833 PGQVDADGNKVTTSQYSADLETRIDKSVT-----PPITGGFSLGASW 874
DB 826 KKNITFEGGNITFGSRKA--VTEIEGNVTNNANVTLLIGSDFDNHQKPLTIKKDVIINS 883
QY 875 KGLSLDADFAIYGVKMINND-----RYFTENAGGIMQLNKXKMLLNATWEDNKETDVP 928
DB 884 GNLTAGGNIVNTAGNLTVESNANFKAITFTFNVGGLFD-----NKGNSNISIA 932
QY 929 KLGQSPQFDTHLEENASFLRLKXKLTYVLPNSLFAGQNVIGGARVYVLMARNL-----L 982
DB 933 KGG-----ARFKDIDNSKMLSITTSSTSYRTIISG-----NITNKGDL 972
QY 983 TVTKYKGFDPPE--AGCNVGVKNQ 1002
DB 973 NITN-EGSDTEMQIGGDVSOKE 993

RESULT 15
AAY34494
ID AAY34494 standard; Protein; 833 AA.
XX AAY34494;
XX
XX 20-MAR-2003 (updated)
DT 25-AUG-1999 (first entry)
XX
XX Porphorymonas gingivalis protein PG35.
XX
XX Porphorymonas gingivalis; PG; periodontal disease; gingivitis;
XX vaccine; antigenic.
XX
XX Porphorymonas gingivalis.
XX
XX WO9929870-A1.
XX
XX 17-JUN-1999.
XX
XX 10-DEC-1998; 98WO-AU01023.
XX
XX 10-DEC-1997; 97AU-0000839.
PR 31-DEC-1997; 97AU-0001182.
PR 30-JAN-1998; 98AU-0001546.
PR 10-MAR-1998; 98AU-0002264.
PR 09-APR-1998; 98AU-0002911.
PR 23-APR-1998; 98AU-0003128.
PR 05-MAY-1998; 98AU-0003338.
PR 22-MAY-1998; 98AU-0003654.
PR 29-JUL-1998; 98AU-0004917.
PR 30-JUL-1998; 98AU-0004963.
PR 04-AUG-1998; 98AU-0005028.
XX
XX (CSLC-) CSL LTD.
PA
XX Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;
XX
XX WPI; 1999-385613/32.
DR N-PSDB; AAX91712.
XX
XX Antigenic Porphorymonas gingivalis peptides for preventing
PT gingivitis
XX
XX Claim 1; Page 474-475; 588pp; English.
PS
XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphorymonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the

Search completed: January 7, 2004, 18:51:14
Job time : 67 secs

Search completed:
Job time : 67 secs

[illegible]

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2004, 18:53:47 ; Search time 27 Seconds
(without alignments)
3622.354 Million cell updates/sec

Title: US-09-581-286A-424
Perfect score: 5298
Sequence: 1 MKRMTLPFLCLLTSGWAMA.....VKNQYPSNQYVAGIQLSF 1017

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	944	17.8	1038	2 JC6027	115K outer membran
2	226.5	4.3	616	2 E83485	probable tonB-depe
3	207	3.9	696	2 D81346	probable iron upta
4	206	3.9	746	2 T46821	siderophore recept
5	206	3.9	746	2 A95420	RhtA Rhizobactin r
6	197.5	3.7	713	2 F82506	probable TonB syst
7	192	3.6	659	2 G91009	colicin I receptor
8	192	3.6	659	2 A85854	hypothetical prote
9	192	3.6	1536	2 A43855	high-molecular-wei
10	190	3.6	599	2 AC3334	metal chelate oute
11	190	3.6	663	1 QREIC	colicin I receptor
12	186	3.5	698	2 AG2219	hypothetical prote
13	185	3.5	3194	2 D71917	toxin-like outer m
14	181	3.4	854	2 AE2082	ferrichrome-iron r
15	180.5	3.4	1268	2 B93789	hemagglutinin/hemo
16	180.5	3.4	1270	2 E85649	hypothetical prote
17	178.5	3.4	819	2 A13197	TonB-dependent rec
18	178.5	3.4	1161	2 S72645	pilV1 protein - Pa
19	177.5	3.4	1029	2 F87369	TonB-dependent rec
20	176.5	3.3	796	2 E87636	ferric aerobactin
21	176	3.3	882	2 AC2082	TonB-dependent rec
22	176	3.3	4919	2 T31105	hypothetical prote
23	175.5	3.3	1902	1 B4858	lactocepin (EC 3.4
24	175	3.3	656	2 E90840	probable TonB depe
25	174	3.3	656	2 E85698	hypothetical prote
26	172.5	3.3	2817	2 B97033	uncharacterized pr
27	172	3.2	863	2 AF2074	ferrichrome-iron r
28	172	3.2	889	2 E87304	TonB-dependent rec
29	170	3.2	742	2 C83061	probable TonB-depe

30 170 3.2 841 2 E71808 probable iron (III
31 170 3.2 1643 2 D71630 outer membrane pro
32 169 3.2 767 2 F64605 iron(III) digitrat
33 168.5 3.2 867 2 AD1856 hypothetical prote
34 168.5 3.2 1579 2 B91290 probable invasin l
35 168.5 3.2 1700 2 G86131 probable invasin Z
36 168.5 3.2 1816 2 F83901 hypothetical prote
37 168 3.2 1649 2 C86822 hypothetical prote
38 167 3.2 1335 2 T17508 glycoprotein Vp260
39 167 3.2 1943 2 B4596 toxin-like outer m
40 166.5 3.1 884 2 H83322 hypothetical prote
41 166.5 3.1 1035 2 C87373 TonB-dependent rec
42 166.5 3.1 1655 2 E97835 hypothetical prote
43 166 3.1 842 2 H64694 iron(III) digitrat
44 165 3.1 1028 2 F87584 TonB-dependent rec
45 164.5 3.1 614 2 AB0935 vitamin B12 recept

ALIGNMENTS

RESULT 1

JC6027 115K outer membrane protein precursor - Bacteroides thetaiotaomicron

N:Alternate names: susC protein
C:Species: Bacteroides thetaiotaomicron
C>Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999
C:Accession: JC6027

R:Reeves, A.R.; D'Elia, J.N.; Frias, J.; Salyers, A.A.

J. Bacteriol. 178, 823-830, 1996

A:Title: A Bacteroides thetaiotaomicron outer membrane protein that is essential for uti

A:Reference number: JC6027; MUID:96146534; PMID:8550519

A:Accession: JC6027

A>Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-1038 <REE>

A:Cross-references: GB:L49338; NID:g1100064; PIDN:AAA95938.1; PID:g1100065
A:Note: it is uncertain whether Met-1, Met-14 or Met-20 is the initiator

C:Comment: This protein is an integral outer membrane protein, and it is essential for

C:Genetics:

A:Gene: susC

C:Keywords: transmembrane protein

F:1-39/Domain: signal sequence #status predicted <SIG>

F:40-1038/Product: 115K outer membrane protein #status predicted <MAT>

F:1028-1038/Domain: transmembrane #status predicted <TM>

Query Match 17.8%; Score 944; DB 2; Length 1038;
Best Local Similarity 27.5%; Pred. No. 3.8e-46;
Matches 315; Conservative 175; Mismatches 407; Indels 248; Gaps 44;

QY 1 MKRMTLPFLCLLTSGW-----AMAQNRVTKGTVISSEDEPLEIGANVVVGGNTTIGAA 54
Db 14 MKKGNFVKVLMLIAGIFLSIDAFQAQITVKG-IVKDTTGPVIGANVVGKTTT-GTI 71
QY 55 TDLQGNFTLSVPANAKMLRVSYGTTTKEVAIANVMKIVLPDPSKVLQVQVVLGYGTQK 114
Db 72 TDFDGNFQLSA-KQGDIIIVSVFIGVQPELPVAAQMNVLKDDTILDEVLVVGQVYK 130
QY 115 LSTVSGSVAKVSSSEKLAEPVANIMDLQGVAGQVQVMTTSGDPTAVASVEIHGTGSLGA 174
Db 131 -NDMTGVSMAIKPDELSKGIITTAQDMLSGIAGSVISNDGTPGGGAQIRIRGSSSLNA 189
QY 175 SSAPLYIVDGMQTSIDVVATM-----NPNDFSMVSLKDSATSIYGAANAGVWFIQ 227
Db 190 SNDPLVIDGLAIDNEGKGMANGLSWMVNPADLETTLVKDASATAIYGSRASNGVIIIT 249
QY 228 TKKGMSRGRITFNASGISQILNTKPLDNMTGDELLDFQVRAGFNGNQTVQKVKDM 287
Db 250 TKKGNGQAPSVTYNGSVFSK---TKRYDVLSDGE----- 283
QY 288 ILAGABDLYGNDYSLKDEYKTLFPVDFNHADMLKALFKTAPTSQGDISFGSGSQGTSY 347
Db 284 YRAVANQLWG--DKLPADLGTG-----NTDQDQIFRTAVSTDHHSVINGGFGKPLPY 333

C;Accession: D81346
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajadream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrel
Nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hyp
A;Reference number: A81250; MUID:20150912; PMID:10688204
A;Accession: D81346
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-696 <PAR>
A;Cross-references: GB:AL139076; GB:AL111168; NID:G6968128; PIDN:CAB73021.1; PID:G696820
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
A;Gene: cfrA; Cj0755
C;Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolo

Query Match	3.9%;	Score 207;	DB 2;	Length 696;
Best Local Similarity	20.1%;	Pred. No. 0.00042;		
Matches 178;	Conservative 113;	Mismatches 280;	Indels 316;	Gaps 48;

QY	101	LEOVVVLGVTGOKLSTVSGSVAKVSEKLAKPVANIMDALQGGVAGQVMMVTSGDPTA	160
DB	25	LDSSIVASGFTQDIKEAPATINVTIKELQSKPYEDVAEI-ADIPGVLDLASKGK---	80
QY	161	VASVEIHGTGSL-----GASSAPLYIVDGMQTSLDVVATMNPDP-----E	201
DB	81	-----TGSYNITWRGITVTLVIDRRQIG--GEVGPNGFNEISNLPPISSIE	130
QY	202	SMSVLKDAASATSYGARAANGVVFIOTKKGWSERGITTFNASYGISQILNTKPLDNMT	261
DB	131	REVIVKGPMT-LYGSEALGGVNNITK--KVSDK-----WETSVSLDALN-----	174
QY	262	GDELLDFQVKAGFWGNQTVQVKMDILAGAEADLYGNYDSLKDEYKTKLPFVDFNHDADW	321
DB	175	--ENKD-----WGNTY-----GTSIVSSGFLMDKLGTL-----	202
QY	322	LKALFKTAPTSQDISFSGSGSQTSYASIGYFDQGMAREPANFKRYSGRLPFSRINE	381
DB	203	---RFREFYRQSQNVETFGS-----GQVQGDQAQSPTKANNFN-----	239
QY	382	WLKVGANLSGATANRRSADYF-----GKYNGSGTFGVLTPRYNPPDVNGDLADVY	435
DB	240	---IGTRIS-YLANDYNTFIDFSRNHYDNKQQLGTITSP-----	278
QY	436	MYGATRPSMTEPYFAKWRPFSSESHQANVGFAQITPI-KGLTLKAQAGVDITNRTSS-	493
DB	279	--GRTPGSIT-----GGVADIMEVDKFTVYLSHEGVYENFISITSL	317
QY	494	--KRMNNPYD---STP--LGERREYARDV---SKSFTNTAEYKFSIDEXHDLTALMG	542
DB	318	QYNRVSDNGREVVGQSTQFPFNGENRDVAEDIIIDTKSV-----IPLQSHILS--VG	368
QY	543	HEY-IEYEGDVTGASKSGFESDKMLLSOGKGTNSLSLPEHRVAEYAVILSPFSRPNYGF	601
DB	369	GEYRLEKMQDKI-ASPTNFDQLLAFAE-----DEYSIKDRLTFTFGARYNH---	415
QY	602	KWYIDFSVRNDOSSRFGSGNNNSAWPYSVCGMFDIYNKFIQESNMLSRLAKWSYCTTCN	661
DB	416	-----HEIFG-NNVSPRAYV-----YNP-----TNELTKGGYSTGFR	448
QY	662	SEIGNYNHOALVTNNYNTEDAMGLSTAGNPDLSWEKOSQFNFLGAAAFNNRLSAEVD	721
DB	449	TPVAN---RLIN-GTYSYSGQG-REPTYGNPDL--KEETSLNYEIAA-IYNNDL-----	494
QY	722	FVYRTNDMLIDVPMPYISGFFSQYQVNGSMK--NTGVDSLKGTIYQNKQWNVYASANF	779
DB	495	FVYSAT-----GFLTNFKDKISQSYNNSEPIPGITGCDADR-----CSRAI	536
QY	780	NYNROBITKLPCLNKYMLPNTGTIWEIGYPNSFYMAEYAGIDKTKQGLWTV--PGQVD	837
DB	537	NHGKVEYKGVELGAGISPLDNL-----NVNFAYTYLDTFVKEAQDRSVIGPQEOS	587
QY	838	ADGNKVTTSQYSADLETRIDKSVTPPIITGFSLGASWKLGLSDADFAYIVGKWMNNDRY	897

[illegible]


```
QY 471 TPIKGLTLKAQGVDTITRTSSKMPNPNPYDSTPLGERRERAYRDVSKSFNTTAE----- 526
Db 312 TOTNRLMANDRGAGILSYRR-----ELRDGFNDEALGATLNAQKEWQ 355
QY 527 -----YKFSIDEKDLTALMGHEY-----IEYEGDVIGASSKGPESDKLMLLS 569
Db 356 QEWLHQFAYGMSVD-----GHDYQRPKSIKRMESGGDDLOADE----- 394
QY 570 QCKTGNLSLPHRVAEYAYLSFFSRFN--YGFDMWYIDFSVRNDQSSRFRGNSRNSAWFY 628
Db 395 -----PPAPAREY-----RFGVYGDNLGLD-----WTL 419
QY 629 SVGCMFDIYNKEIQESNWLSDLRLKMSYGTGTSNIGYNNHQAULTVNNYTEDAMGLSIS 688
Db 420 AAGLFDAQ-----KLSPKNT--DRIHGYK-----VVTMGS-SEWSPASIS 458
QY 689 TAGNPDLSEWKSQNFGLAAGAF-----NNRLSAB----- 719
Db 459 YQWHEP--WNTYLSYNNHGFRAPSYDKAYGASHSFVPLTPFTIKNNKLRABETSDFELG 516
QY 720 -----VDYVYR--TTNDMLIDVPM-----PYISGFFSQYQNVGSMKNTGVDLG----- 760
Db 517 SKYDNGQOTFYVAVFYISFDNFIDYKQVYDNATGSLVQIQYQNIAGVKTGYGAEMSVMERL 576
QY 761 -----LKGTY-----YQKDMNVYASAF--NYNRQBIT 787
Db 577 DDRWSVENKLYGVGDKGQENQVYRTLPLEGSQLNYQRRWDAYSRLNWSAMSRRVPTC 636
QY 788 KLFGGLNKYMLPNTGTI-WEIG 808
Db 637 TTEQKTECATTTGVNSWDIG 658

RESULT 7
G91009
colicin I receptor precursor [imported] - Escherichia coli (strain O157:H7, substrain RI
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: G91009
R:Havashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: G91009
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-659 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA036470.1; PID:g13362516; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
C:Gene: ECa3047
C:Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolo

Query Match 3.6%; Score 192; DB 2; Length 659;
Best Local Similarity 20.2%; Pred. No. 0.0028;
Matches 178; Conservative 103; Mismatches 272; Indels 330; Gaps 50;

QY 92 IVLDPDSKVLQVVLVGTGQKLTSGVSGVAKVSEKLAEPVANIMDALQGQVAGMNV 151
Db 24 LAVDDG---ETMVYVYASSVEQNLKDPASISVITQEDLQRPQNLKDLVKL-EVPGVQ- 78
QY 152 MTTSGDPTAVASVEHTGTSGLGASSAPLYIVDMQTSLDVVATMNPDPF-----E 201
Db 79 LNEGDNR--KGVSTIRGLD-----SSYTLILVDGKRVN--SRNAVFRHNDPDLNWIIPVDSIE 131
QY 202 SMSVLKASATSIYGARAANGVVFQTKKGMSEGRITFNASYGISQLNTKPLDNMT 261
Db 132 RIEVVR-GPMSSLYGSDALGGVNNITTKIGQKWSGTVVDT----- 172
QY 262 GDELLDFQVAGFWGNQTVQVKMDMILAGBDLVGNVDSLKDEYKTKLPFVDFNHDADW 321
Db 173 -----TIQHRD-----RGDTYNG----- 186
```

```
QY 322 LKALFKTAPTSGQDISFSGSGQTSYASIGYFD-----QEGMAREPANFKRYSGRLNFES 377
Db 187 --QFPTSGLIDGVL-----GMKAYGSLAKREKDDPQNSTTTDTGTGTPIEG---FSS 234
QY 378 RINELWLVKANLSGAIANRRSADYFGKYMGSGTGVLTMPRYNPPFDVNGDLADVYMY 437
Db 235 R-----DGNVEFAWTFNQHDF-----TAG-----Y 255
QY 438 GATRPSMTEPYFAKMRPFSSSHQANVNGFAQITPIKGLTLKAQAGVDITNTRTSSKMP 497
Db 256 GFDRQDRDSDLDKNR--LERQYVSVMGRWDY---GTSELKYYGEKVEN----- 301
QY 498 NNPYDSTPLGERRERAYRDVSKSFNTTAEYKFSIDEKDLTALMGHEYIEYEGDVIGASS 557
Db 302 KNPGNSSPI-----TSESNTVDGKYTL-----PLTAI--NQFLTVGGE----- 337
QY 558 KGFESDKL--MLLSQKGTGNLSLPHRVAEYAYLSFFSRFNYPGDKM-----MYIDFSV 610
Db 338 --WRHDKLSDAVNLTTGGTSKTS-----ASQVAL-----FVDEWRIPEPLALTTGV 382
QY 611 RNDQSSRFGSNRSNAWFYSGVMFDIYNKFTQESNWLSDLR-----LKMSYGTGTSNIGN 666
Db 383 RNDHETTYGER-----WSPRAYLVVNATDVTTVKGGWATAFRAPSLQLQSDPMTSNCRG- 437
QY 667 YNHOALVTNNYTEDAMGLSISTAGNPDLSEWKSQNFGL-----ACKIVGSPDLKPETSSEWELGYMGEGLGVESVTVF 478
Db 438 -----ACKIVGSPDLKPETSSEWELGYMGEGLGVESVTVF 478
QY 713 NNRLSAEVDFYVRTNDMLIDVPMPIYISGFFS-----OYQNVGSMKNTGVDLSLKGTIY 766
Db 479 RNDVKDRIS--ISRTSDVNAAPGYQNFVGTANGVPVPSYNNVANKARIQGVETELK--IP 534
QY 767 QNKDMNVYASAFNFIN--RQETIKLFPGLNKYM--LP-----NTGTI--WEIGYPNPSFYMAEYA 819
Db 535 FNDEWKL--SLNYTYNDGRDVSN---GENKPLSOLPHTANGTLDM----- 575
QY 820 GIDKKTGKQLM--YVPGQVADAGDNKVTQYSADLETRIDKSVTPPIITGGFSL---GASW 874
Db 576 ---KPLALEDSFVYSGHY-----TGQKRD-----SATAKTGGYTIWNTGAAM 617
QY 875 -----KGLSLDADFAYIVGKWMINND--RYF 898
Db 618 QVTQDKVLRAGVNLGDKLSD--DYSY-----NEDGRYVF 652

RESULT 8
A85854
hypothetical protein cirA [imported] - Escherichia coli (strain O157:H7, substrain EDL93
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: A85854
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: A85854
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-659 <STO>
A:Cross-references: GB:AS005174; NID:g12516475; PIDN:AAG57293.1; GSPDB:GN00145; UWGP:234
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
C:Gene: cirA
C:Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolo

Query Match 3.6%; Score 192; DB 2; Length 659;
Best Local Similarity 20.2%; Pred. No. 0.0028;
Matches 178; Conservative 103; Mismatches 272; Indels 330; Gaps 50;

QY 92 IVLDPDSKVLQVVLVGTGQKLTSGVSGVAKVSEKLAEPVANIMDALQGQVAGMNV 151
Db 24 LAVDDG---ETMVYVYASSVEQNLKDPASISVITQEDLQRPQNLKDLVKL-EVPGVQ- 78
QY 152 MTTSGDPTAVASVEHTGTSGLGASSAPLYIVDMQTSLDVVATMNPDPF-----E 201
Db 79 LNEGDNR--KGVSTIRGLD-----SSYTLILVDGKRVN--SRNAVFRHNDPDLNWIIPVDSIE 131
QY 202 SMSVLKASATSIYGARAANGVVFQTKKGMSEGRITFNASYGISQLNTKPLDNMT 261
Db 132 RIEVVR-GPMSSLYGSDALGGVNNITTKIGQKWSGTVVDT----- 172
QY 262 GDELLDFQVAGFWGNQTVQVKMDMILAGBDLVGNVDSLKDEYKTKLPFVDFNHDADW 321
Db 173 -----TIQHRD-----RGDTYNG----- 186
```

Db 24 LAVDDG---ETMVTASSVEQLKDPASISVITQEDLQKFEVQNLKDKVLK-EVFGVQ- 78
Qy 152 MTTSGDPTAVASVEIHGTGSLGASSAPLYIVDGMQTSLDVVATMNPDR-----E 201
Db 79 LTVGNR--KGVISGLD-----SSYTLIVDGRVN-SRNAVFRINDPDLNIPVDSIE 131
Qy 202 SMSVLKASATSIYARAANGVVFVIOTKGKMSGRITFNASYGISQILNTPKLDNMWT 261
Db 132 RIEVVR-GPMSSLYGSDALGVVNIITKKIGQKSGTVTVDT----- 172
Qy 262 GDELLDFQVKGAGFNWQNTQVKQKMLAGAEGLYGNYSKDEYKTLPPVDFNHDADW 321
Db 173 -----TIQEHDR-----RGDTYNG----- 186
Qy 322 LKALEKTAPTSQGDISFGSGSQTSYVASYGYFD-----QSGMAREPANFKYSGRLNPFES 377
Db 187 --OFFTSGPLIDGL-----GNKAYGSLAKREKDDPQNSTTDTGTETPRIEG-----FSS 234
Qy 378 RINEWLKVGANLSAGAIANRRSADYFGKYMGSGTFFGLTMPRYNPPDVNGDLADVYMY 437
Db 235 R-----DGNVEFAWTPNQHDF-----TAG-----Y 255
Qy 438 GATRPSTPEYFAKMPFFSSESHOANVNGPAQITPIKGLTKAQAGVDITNRTSSKRM 497
Db 256 GFDRQDRDSLDKNR-LERQNTYSVSHNGRDY-----GTSELKYGEKVEN----- 301
Qy 498 NNPDSTPLGERERAYRVSKFTTAETAEYKFSIDKHDLTALMGHEIYEYEGDVIGASS 557
Db 302 KNFGNSSPI-----TGESNTVDGKYTL-----PLTAI--NQFLTVGGE----- 337
Qy 558 KGPESDKL--MLLSQKGTGNSLSLPEHRVAEYAYLSFFSRFNYGFDKW-----MYIDFSV 610
Db 338 --WRHDKLSDAVNLGTGTSKTS-----ASQYAL-----FVEDEWRIFEPLALTGV 382
Qy 611 RNDQSRFGSNRRSAPFYSVCGMFDYINKFIQSNMLSDLR-----LKMSVGTGTSNBEIGN 666
Db 383 RMDHETGYEH-----WSPRAYLYNATDVTIVGGMWATAFAKSLQLSDMTSNCRG- 437
Qy 667 YNHQALVTNNYTTEDAMGLSISITAGNPDLSEKQSFNEGL-----AAGAF 712
Db 438 -----ACKIVGSPDLKPETSESWEGLYMGEGWLEGVSSVTVP 478
Qy 713 NNRLSAEVDVYRTNMLIDVMPYISGFFS-----QYQVGSMMKNTGVDLSLKGTY 766
Db 479 RNDVKDRIS--ISRTSDVNAAPGVQVFGTANGVPFVSYYNKNKARIQGVETELK--IP 534
Qy 767 QNKDMNYYASANFNYN-ROBITKLFFGLNKM--LP-----NTGTI-WEIGYFNSFYMAEYA 819
Db 535 FNDEWKL--SLNTYNDGRVSN--GENKPLSDLPHTANGTLW----- 575
Qy 820 GIDKKTGKQLW--YVPGQVDADGNKVTTSQYSADLETRIDKSVTPPTGGSFL--CASW 874
Db 576 ---KPLALEDWSPYVSGHY-----TGQKRAD-----SATAKTPGGYTIWNTGA 617
Qy 875 -----KGLSLDADFAIYVKGWMINND--RYF 898
Db 618 QVTKDKVLRAGVLNGLDKOLSRD-DYSY-----NEDGERYF 652
RESULT 9
A43855
high-molecular-weight surface-exposed protein HMW1 - Haemophilus influenzae
C:Species: Haemophilus influenzae
C:Date: 10-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
R:Accession: A43855
R:Barankamp, S.J.; Leininger, E.
Infect. Immun. 60, 1302-1313, 1992
A:Title: Cloning, expression, and DNA sequence analysis of genes encoding nontypeable Haemophilus pertussis.
A:Reference number: A43855; PMID:92192797; PMID:1548058
A:Accession: A43855
A:Status: preliminary
A:Molecule type: DNA

A;Residues: 1-1536 <BAR>
A;Cross-references: GB:M84616; NID:9475770; PIDN:AAA20527.1; PID:9475771
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:89235, NCBIIP:89239)

Query Match 3.6%; Score 192; DB 2; Length 1536;
Best Local Similarity 18.6%; Pred. No. 0.011;
Matches 205; Conservative 151; Mismatches 378; Indels 368; Gaps 51;

Qy 25 VGTVVISSEDNREPLICANVVVG-----NTTGAATDLDGNFTLSVPAKMLRVSYSGM 79
Db 136 LKG-ILDSNGQVFLNPNGITTKDAIINTNGFTASTLD-----ISNENIKARNPTE 187
Qy 80 TTKEVAIANVMK---IVLDPD-----SKVLEQVVVLGTCQKLSVTSVSGSVAKTSSEK 129
Db 188 QYKDALAIIVNHGLITVGKDSVNLIGCKVKNVGI-----SVNGGSIALLAGOK 238
Qy 130 LAEKPVAN-----IMDALQGVAGMQVMTSGDPTAVASVEIHGTGSLGASSAPLYVDG 184
Db 239 ITISDIINPTITVSTIAAPENAVNLGDIPAKGGNINVRATIRNQKLSAD----- 289
Qy 185 MQTSLDVVATMPNDPESMSVLKASATSIYGARAA-----NGVVFIOTKGKMSERGIT 240
Db 290 -----SVSKDKSGNIVLSAKEGEAIGGVISAQNOQAK---GGKL- 326
Qy 241 FNASYGISQILNTPKLDNMVTGDE-----LLDFQVKAG---FWGNNQTVQVKDMILA 290
Db 327 -----MITGDKVTLTKGAVIDLSGEGGETYLGDERGEGKNGIQLA 368
Qy 291 GAEDLYGNYD---SLKDEYKTLF-----PVDFNHDADMLKALFKTAPTSQGISPSGGS 342
Db 369 KKTSLKSGSTINVSKEGKGRAIVMGDIALIDGNINAQ-----GSGDIKATGGF 417
Qy 343 QGTS-----YYASIGYFDQEGMAREPANFK---RYSGLNLFESRINELKVGANLSGAI 395
Db 418 VETSGHDLFIKNAIVDAKEWLLDPNVSINAEIAGRSN-TSEDDYTGSGNSASPTKRN 476
Qy 396 RRSADYFG---KYVWGSSTFGVLTMPRYNPPDVNGDLADVYMYGATRPSMTEPYFAKM 452
Db 477 KEKTLTNTLSILKKGTFVNI-----ANQ 503
Qy 453 RPFSSSHOANVNGPAQITPIKGLTL-----KQAGVDITNRTSSKRMNPNVDSPTLGE 508
Db 504 RIYVNSSINLS-NG-----SLTLWSRSGSGGVEINNDIT- 540
Qy 509 RRERA-----YRDVSKSFTNTAEYKFSIDKHDLTALMG-HEYIEYEGDVIGASSKG 559
Db 541 DTRGANLTIYSGGVVDVHKNI SLGAGQGNINITAKQDIAFEKGSNQVITQGTITSGNQK 600
Qy 560 FESDKMLLSQGTGNSLSLPEHRVAEYAYLGFSSRFNYGFDKMMYIDFSVRNDQSSRF 619
Db 601 FRFNVSL---NGTSGLOFTTKRTNKYAITNKF-----G 633
Qy 620 SNRSAMFYSVGGMFDI-----YKFKIQESNW-LSDRLKMSYGTGTGNSGTGNYN 668
Db 634 TLN-----ISGKVNISMVLPKNESGYDKFGRTYWNLTSLN-----SESSEFN 677
Qy 669 HQALVTNNYTTEDAMGLSISITAGNPDLSEKQSFNEGLAAGAFNNRLSNEVDVYRTTN 728
Db 678 -----LTIDSRGSDSAGTLTPQYNLNGISFNKOTTFNV-----ERNARVNF----- 718
Qy 729 DMLIDVPMF-----YISGFFSOYQ-NVGSMMKNTGVDLSL-----KGTIYQNKDMV 773
Db 719 -----DIKAFIGNKSSLYNAYFNGNISVSGGSDVFTLLASSSNVQTFGVVINSKYFNV 774
Qy 774 YASAFNENYRQBITKLFFGLNKMYPN-TGTIWEIGYFNSFYMAEYAGIDKKTGKOLWYV 832
Db 775 STGSLRFPKTSSTKGTGFSIEKDLTLNATG-----GNITLLQVEGTDGMIGKGI-VA 825
Qy 833 PQGVADAGNKVTSQYSADLETRIDKSVT-----PITGGSFLGASW 874
Db 826 KKNITPEGGNITFGSRKA--VTEIEGNVTNNANVTLIGSDFDNHOKPLTIKKVVIINS 883

QY 875 KGLSLDADFAIVGKMINND-----RYFTENAGGLMQLNKKDKMLLNATWEDNKETDVP 928
Db 884 GNLTAGGINIVNAGLITVENANFAITNFTFNVGGLFD-----NKGNSNISIA 932
QY 929 KLGQSPQPDTHLENASFLRLKQLKLTIVLPNSLPAGQNVIGGARVYLWARNL-----L 982
Db 933 KGG-----ARFKDINSKSLITNSSSTYRIISG-----NITKNGDL 972
QY 983 TVTKYKGFDPPE--AGGNVKGKQ 1002
Db 973 NITN-EGSDTEMOIGGDVSQKE 993

RESULT 10
AC3334
metal chelate outer membrane receptor [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 23-Dec-2002
C:Accession: AC3334
R:Delvecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujar, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltzman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AC3334
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-599 <KUR>
A:Cross-references: GB:AE008917; PIDN:AA151838.1; PID:gl7982585; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BME10657
A:Map position: 1
C:Superfamily: vitamin B12 receptor; tonB-dependent receptor amino-terminal homology; to

Query Match 3.6%; Score 190; DB 2; Length 599;
Best Local Similarity 18.9%; Pred. No. 0.0031;
Matches 155; Conservative 114; Mismatches 275; Indels 274; Gaps 37;

QY 97 DSKVLEQVVLGYGTGQKLSLVGSGVAKVSSEKLAEPVANIMDALQGVAGMQVMTSG 156
Db 10 DGVILDTIVTPLRASSLQSTSVSVIDAADI-ERSAAPDLQLLQTYGSGVKTNGG 68
QY 157 DPTAVASVEIHGTGSLGASSAPLYIVDQMOTS-----LDVVATMPNDFESMVLKQASA 211
Db 69 QG---SSADIYMRGM--SSKQTVLVNGVVRTASATSGSTALANIPLTSIERIEIARGAH- 122
QY 212 TSIYGARANGVVFQTKGKMSERGRITFNASYGISQILNPKLDNMGTG-----DEL 265
Db 123 SSQYGADAIGGVINIITKGGACGERACWCGSVSTGVSH-----PWGGYASGSLQGRSSDG 177
QY 266 LDFOVKAGFWGNQTVQVKDMILAGAEGLYGNVDSLKDEYKTLFPVDENHDADWLKAL 325
Db 178 IDYAVGAALTGTQ-----GYDFTT-----PEAFGHEPD----- 205
QY 326 FKTAPTSOGDISFSGSGSGTSYASIGYFDQEGMAREPANFRYSGRNLNFE-SRINELWK 384
Db 206 -----DDGFLQGS-----FNFALSXKDFDNGK 226
QY 385 VGANLSGAIANRRSADYFGKYMGGTGFVLTMPRYNPFVNGDLADVYMYGATRPSM 444
Db 227 IYAD--GLFSRGRN-----QYDADAPAF 247
QY 445 TEPYFAKMRPFSSSE-----SHQNVNGFAQITPIKGLTLKAQAGVDITNTRTSKGMENN 499
Db 248 NE---ADSTAPTKGVGTIDHTADWSS-----TVFSTGID--NSNFRKGIEGS 292
QY 500 PYDSTPLGERERARVDVSKFTNTAAYKFSIDEKHLDTALMGHBYIEYEGDVIGASSKG 559
Db 293 DWFEI---RYGVFPASTEKSF-----TGKVS-----VVTGGVEA 325
QY 560 FESDKMLLSQCKTGNLSLPEHRVAEYAYLSFFSRFNYGDKMYIDFSVRNQSSRRFP 619

Db 326 YREKINTTIDYETGRDLA-----AVFGQYSLEYDA-LRFPDGGIRYDHNGQPG 372
QY 620 SNNSANFYSVGGCFDIYNKEIQBSNWLSDLRLLKMSYGTTCGNSEIGNYNHQAIVTVNNYT 679
Db 373 ---NVITYNLGASVEI-----LPDLVRSYATGFRAPTFN----- 405
QY 680 EDAMGLSISTAGNPDLSEKQSQNFGLAAGAFNNRLSAEVDYVYRTTNDMLIDVPMPIY 739
Db 406 ---ELYYPGFANPDQLQPEKRSVEVGLNWQA-TASTSLDMALYQTRLSDAIMSTAPSYI 460
QY 740 SGFFSQYQNVGSMKNTGVDLSLKGTYIYONKDNV-----YA 775
Db 461 P-----YNIASAKVTGLEATLSHSF--NEQWGIKGMVDLKRPFVDEDSNDLPYRERPKA 512
QY 776 SANFNYNRQE-----ITKLPFGLNKMLPNTGTIWEIG-YPNSEFYMAEYAGIDKTKGQLW 830
Db 513 AAEVNFPEVKLDLTARVLYGGSRY--TNKNTKLGIVYVTFADFVALYS-IDKQS--QLK 567
QY 831 YVPGQVDADGNKVTTSQYSADLETRIDKSVTPPTGGF 868
Db 568 PSVENI-PDKDYETSSGYVAP-----GRITITGLTRNF 599

RESULT 11
QREIC
colicin I receptor precursor - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C>Date: 30-Sep-1990 #sequence_revision 05-Dec-1997 #text_change 01-Mar-2002
C:Accession: B64984; A32056; A33868; A28377; A35408; S24561
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: B64984
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-663 <BLAT>
A:Cross-references: GB:AE000304; GB:U00096; NID:gl788470; PIDN:AACT5216.1; PID:gl788478,
A:Experimental source: strain K-12, substrain MGL655
R:Nau, C.D.; Konisky, J.
J. Bacteriol. 171, 1041-1047, 1989
A:Title: Evolutionary relationship between the TonB-dependent outer membrane transport p
A:Reference number: A32056; MUID:89123100; PMID:2644220
A:Accession: A32056
A:Molecule type: DNA
A:Residues: 1-146, RCARRCEVHHQ', 159, 'NRSEMV', 166, 'YRPRYHHSGTSRR', 182-527, 'N', 529-613,
A:Note: the authors translated the codon AAT for residue 528 as Ile
R:Nau, C.D.; Konisky, J.
J. Bacteriol. 171, 4530, 1989
A:Reference number: A33868
A:Contents: corrections
A:Accession: A33868
A:Molecule type: DNA
A:Residues: 144-184; 610-663 <NA2>
R:Griggs, D.W.; Tharp, B.B.; Konisky, J.
J. Bacteriol. 169, 5343-5352, 1987
A:Title: Cloning and promoter identification of the iron-regulated cir gene of Escherich
A:Reference number: A28377; MUID:88058737; PMID:3316180
A:Accession: A28377
A:Molecule type: DNA
A:Residues: 1-59 <GRI>
A:Cross-references: GB:M19295; NID:gl45545; PIDN:AAA23581.1; PID:g551794
A:Note: residues 26-45 were confirmed by protein sequencing
R:Steffes, C.; Ellis, J.; Wu, J.; Rosen, B.P.
J. Bacteriol. 174, 3242-3249, 1992
A:Title: The lypS gene encodes the lysine-specific permease.
A:Reference number: A41871; MUID:92250419; PMID:1315732
A:Accession: C41871
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-96, 'D', 98-125 <STE>
A:Cross-references: GB:M89774; NID:g466776; PIDN:AAA17054.1; PID:g466779; EMBL:X65029

R;Griggs, D.W.; Kafka, K.; Nau, C.D.; Konisky, J.
J. Bacteriol. 172, 3529-3533, 1990
A;Title: Activation of expression of the Escherichia coli cir gene by an iron-independent
A;Reference number: A35408; MUID:90264362; PMID:2160948
A;Accession: A35408
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-21 <GR2>
C;Comment: This outer membrane receptor for colicins Ia and Ib is regulated by both cell
membrane transport proteins.
C;Genetics:
A;Gene: cir; cirA; feuA
A;Map position: 43 min
C;Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolo
P;1-25/Domain: signal sequence #status predicted <SIG>
P;46-624/Product: colicin I receptor #status experimental <WAT>
P;66-211/Domain: tonB-dependent receptor amino-terminal homology <TNN>
P;365-663/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>
Query Match 3.6%; Score 190; DB 1; Length 663;
Best Local Similarity 20.3%; Pred. No. 0.0037;
Matches 180; Conservative 102; Mismatches 270; Indels 336; Gaps 51;
QY 92 IVLDPDSKVLEQVVVLGYGTGQKLVSTVSGVAKVSEKLAEPVANIMDALQGVAGMQV 151
DB 24 LAVDDG---ETWVTVASSVEQNLKADAPASISVITQEDLQKVPQNLKVLK-EVPGVQ- 78
QY 152 MTSGDPTAVASVEIHGTGSLGASSAPLYIVDGQTSLDVVATWPNDF-----E 201
DB 79 LTNEGDNR--KGVSIKGLD---SSYTLILVDGKRVN-SRNAVRHNDPLNIPVDSIE 131
QY 202 SMSVLKASATSIYGARAANGVVFIOTKKMSERGRITNASYVGIQLLNTPLDNMT 261
DB 132 RIEVVR-GPMSSLYSGDALGVNIIITKIKQKWSGTVTVDT----- 172
QY 262 GDELLDPQVKAFGWGNQTVQVKDMILAGAEDLYGNVDSLKDEYKTLFPVDFNHDAW 321
DB 173 -----TIQHRD-----RGDTYNG----- 186
QY 322 LKALFKTAPTSQGDISFSGSGQTSYYASGYDF-----QGMAREPANFKRYSGRLNFS 377
DB 187 --QPTSGPLIDGVL-----GMKAYGSLAKREKDPQNSTTTDTGETPRIEG---FSS 234
QY 378 RINELWKVGNLSAIGANRSDADYFGKYMGSTFGVLTPRYNPPDVNDLADVYMY 437
DB 235 R-----DGNVFAPTPNQHDF-----TAG-----Y 255
QY 438 GATRPSTPEYFAKMRPFSSSEHQANVNGFAQITPIKGLTLKAQAGVDITNTRTSKRMP 497
DB 256 GFDQRDRSDSLKNR-LERQNVSVSHGRWDY-----GTSELKYGEKVEN----- 301
QY 498 NNPDVSTPLGERRERAVRVSKFTNTAEYKFSIDEKHDLTALMGHEYIEYEGDVTGASS 557
DB 302 KNFGNSPI-----TSSENTVDGKYTL-----PLTAI--NOFLTVGGE----- 337
QY 558 KGFESDKL--MLLSQGTGNSLSLPEHRVAEYAYLSFPFRNYGFPDKW-----MYIDFSV 610
DB 338 --WRHDKLSDAVNLGTGTSKTS-----ASQYAL-----FVEDEWRIEFPLATTGV 382
QY 611 RNDQSRFRGNNRSAMFYSGVGMFDIYNKFIQESNMLSDLR-----LKMSYGTGNSBEIGN 666
DB 383 RMDHETVYGEH-----MSPRAYLYNATDTTVVKGWATAPKAPSLQLSPDWTNSCRG- 437
QY 667 YNHQALVYNNYTEDAMGLSISTAGNPDLSWEKQSQNFGL-----AAGAF 712
DB 438 -----ACKIVGSPDLKPETSESWEGLGYMGEGMLEGVESVTVF 478
QY 713 NNRLSAEVDYVRTNDMLDVPMPYISGF-----FSQYQNVGSKMTGYDLSL 761
DB 479 RNDVKORIS--ISRTSDVNAAPGYQNVFQFETGANGRRIPVFS--YNNVNKARIQGVETEL 535
QY 762 KGTIYQNKDWNVYASANFNYN-RQEITKLPFGLNKM--LP---NTGTI-WEIGYPNSFY 814

Db 536 K--IPNDEWKL--SINYTYNDRGVSN--GENKPLSDLPFHTANGTLDW----- 579
QY 815 MAEYAGIDKKTGKQLW--VPGQVDADGNKVTTSQYSADLETRIDKSVTPPTTGGFSL-- 870
Db 580 -----KPLALEDWSFYVSGHY-----TGQKRAD-----SATAKTPGGYTIWN 616
QY 871 -GASW-----KGLSLDADFAFYVIGKWMINND--RYF 898
Db 617 TGRAWQVTKDKVRAGVNLGDKLSRD-DYSY-----NEDGRYF 656
RESULT 12
AG2219
hypothetical protein all3310 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AG2219
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AG2219
A;Molecule type: DNA
A;Residues: 1-698 <KUR>
A;Cross-references: GB:BA000019; PIDN:BA075009.1; PID:gi17132405; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all3310
Query Match 3.5%; Score 186; DB 2; Length 698;
Best Local Similarity 18.5%; Pred. No. 0.0066;
Matches 165; Conservative 130; Mismatches 289; Indels 308; Gaps 39;
QY 37 PLIGANVVV---VGNTTIGAATDLDGNFTLSVP-ANAKMLRVSYSGMTTKEVAIANVMKI 92
DB 20 PALGSTETITQVVDNSDIPYLSEI-----LPTTSABELLTQSTPG----- 59
QY 93 VLDPDSKVLEQVVVLGYGTGQKLVSTVSGVAKVSEKLAEK---PVAN---IMDALQGV 146
DB 60 DLNPECEPSEQ-----PEIEETSDADITIEAIAEPETLPASTPTTYVIDQBEIQ 110
QY 147 AGWQVMTTSGDPTAVASV-----EIH-GTGLSGAS-SAPLYIVDGQTS 198
DB 111 QG-----ATSVADVLKRMPPGFAINDVGHGADIHTGTYYRGASINOSVFLINGRRIN 161
QY 189 LDVVATMNPDPRESMSV-----LKDASATSIYGARAANGVVFIOTKKMSERGRITF 241
DB 162 NDVNTYHGTATDLNSIIPVESIERVELSSGVTSAIYGSASAGGVVNIITKG--YPOPKLTS 219
QY 242 NASYGISQLINTKPLDNMTGDELLDFQVKAQWGNNTVQVKDMILAGAEDLYGNYS 301
DB 220 SLEFGSLNLNQFSGVGAATYNFSPERYVDNRYRP-----VGAA----- 264
QY 302 LKDEYKTLFPVDVFNHDAWLKALFKTAPTSQGDISFSGSGQTSYYASIGYDFQEGMAR 361
DB 265 -----NRD-----SQGLSNADTSTST-YFGNIG----- 287
QY 362 EPANFKRYSGRLNFESRINEMLKVGANLSAIGANRSDYFG----- 403
DB 288 -----LDLQORNLSLDI-----TKLSSRGLVYFGPQLDRDLHDGLNIGLS 331
QY 404 -KYMGSGTGFVLTMPRYNPDVNGDLADVYMYGATRPSTPEYFAKMRPFSSSHQA 462
DB 332 WKTRLGNGNSNLITTFGYN-----QNYFSTYGTPTVFAGRE--FYRTGLVDTQQTG 381
QY 463 NVNGFAQITPIKGLTLKAQAGVDITNTRTSKMPNPNVDSTPLGERRERAYRDVSKSFT 522
DB 382 RIDHWRISP-----NNKLRWGLDKNT----- 404


```

Db 706 TLQSSSGTMDNNATGFTASDKNLSLEVNSLT---NYGWSIGKGDVD----- 751
QY 565 LMLLSQGTGNSLSLPEHRVAEYAYLSFFSRENYGDKWYIDFSVRND---QSSRFQSN 621
Db 752 -VTVANGNLYNRNTIAAEKGLDIAALN-----GIENWK--DISAGGDLTWTNRHVTN 801
QY 622 NRSANFYSGGPFDIYNKFQESNWLSDLRLKMSYGTTCNSEIGNYNHOA----- 671
Db 802 NSNSNMVGQIYINAVNDINNRGNIVSDADLNV--TKGN--LYNYLYMVGYGDIALSAN 857
QY 672 LVTVNNYTEDAMG-LSISTAGNPDLSEWKQSQNFGLAAGAFN-NRLSAEVDFFVYRTND 729
Db 858 SVANNNAITEATGDLIIDSQNGVGNRRGNLHALNGVLSVKGNNLNDNGEIRGYG----- 912
QY 730 MLIDVPMFYISGFFSQYQNVGSMKNTGVDLSLKGTYIQNKDWNVYASAFN-----YN 782
Db 913 ---DVTLA-LTCNYDSYK--GSLTSETGDTVLTANIVDNA-YGLIAGENVSVDAKSTIYN 965
QY 783 QEITKLPFGLNKYMLPWTGTIWEIGYPNSFYMAEYA--GIDKTKGQLWYVPGQVDADG 840
Db 966 NTAL-----IAANKKLVINAGGNLENRDNFNLRNNGALFGITDNVG-----GIVGKEG 1014
QY 841 NKVTTSQYSADLETRIDKSVTPPTGGFSLGASWKGLSLDADFAIVGKWMINNDRYTE 900
Db 1015 VTLSAQNVYNNSSIAENGPLNLSRGFLDNRALLSSGAD-AIIRAAGTFYNNYATY 1073
QY 901 NAGGL-----MQLNKOKMLLNAWTEKNKEDVPEKLGQSQFDTHLLENASFLRLKNLKLTY 956
Db 1074 SAGNLDVYAASLNNAS---DGRLEDNTATGV-----IASDKNLDLS- 1111
QY 957 VLPNSL-----PAGQ-----NVIGGA---RVYLMARNLLTVTKYKGFDPGAGNVGKNQY 1003
Db 1112 -VDNSVTNYGWSIGRGDVHFVNLKGTLYNRNAIAADNALTINALNGVE----- 1158
QY 1004 PNSKOYVAGIQLS 1016
Db 1159 -NFKDIVAGTALT 1170

```

Search completed: January 7, 2004, 19:04:58
Job time : 34 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2004, 18:51:42 ; Search time 18 Seconds
(without alignments)
2657.009 Million cell updates/sec

Title: US-09-581-286A-424
Perfect score: 5298
Sequence: 1 MKRWTLFLCLLTLSIGWAMA.....VGKNQPNKQYVAGIQLSP 1017

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	206	3.9	746	1 RHTA RHIME
2	190	3.6	663	1 CIRA_ECOLI
3	175.5	3.3	1902	1 P2P_LACPA
4	170	3.2	1643	1 OMPB_RICPR
5	166.5	3.1	1655	1 OMPB_RICCN
6	166.5	3.1	1953	1 BIGA_SALTY
7	166	3.1	2003	1 YDBA_ECOLI
8	161.5	3.0	614	1 BTUB_SALTY
9	161.5	3.0	944	1 LBPA_NEIMA
10	161.5	3.0	2249	1 OMPA_RICRI
11	160	3.0	1061	1 OAR_MYXXA
12	160	3.0	2021	1 OMPA_RICCN
13	159	3.0	652	1 IRGA_VIBCH
14	158.5	3.0	2710	1 TOXA_CLODI
15	157	3.0	723	1 Y262_HAEIN
16	154.5	2.9	1140	1 YD76_MYCPN
17	154.5	2.9	1902	1 P1P_LACLC
18	153.5	2.9	987	1 YD54_METUA
19	153	2.9	905	1 HXAL_HAEIN
20	152	2.9	869	1 CFAC_ECOLI
21	152	2.9	1654	1 OMPB_RICRI
22	151.5	2.9	1902	1 P3P_LACLC
23	150	2.8	673	1 FYUA_YEREN
24	149.5	2.8	700	1 YNCD_ECOLI
25	149	2.8	673	1 FYUA_YERPE
26	148	2.8	1256	1 MRP_STRSU
27	148	2.8	1286	1 AIDA_ECOLI
28	148	2.8	1645	1 OMPB_RICTY
29	148	2.8	1656	1 OMPB_RICJA
30	147.5	2.8	1902	1 P2P_LACLC
31	146.5	2.8	810	1 HPUB_NEIMA
32	145.5	2.7	614	1 BTUB_ECOLI
33	145.5	2.7	828	1 PMFC_PROMI

34 145 2.7 725 1 HXC2_HAEIN P45357 haemophilus
35 144.5 2.7 784 1 OSTA_ECOLI Q8CWE6 escherichia
36 143.5 2.7 784 1 OSTA_ECOLI P31554 escherichia
37 142 2.7 1409 1 HAP1_HAEIN P44596 haemophilus
38 142 2.7 3305 1 APLP_MANSE Q25490 manduca sex
39 141.5 2.7 784 1 OSTA_ECOS7 Q8XAL3 escherichia
40 140.5 2.7 1577 1 HUYA_PROMI P16466 proteus mir
41 140 2.6 937 1 CS32_ECOLI P15484 escherichia
42 139 2.6 790 1 YDDB_ECOLI P31827 escherichia
43 139 2.6 943 1 LBPA_NEIMB Q06379 neisseria m
44 139 2.6 1314 1 TETX_CLOTE P04958 clostridium
45 139 2.6 2358 1 MOKD_SCHPO Q9Y719 schizosacch

ALIGNMENTS

RESULT 1
RHTA RHIME
ID RHTA RHIME STANDARD; PRT; 746 AA.
AC Q923Q5;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Rhizobactin receptor precursor (TonB-dependent siderophore receptor
DE rhtA)
GN RHTA OR RA1265 OR SMA2414.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Plasmid pSyma (megaplasmid 1).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RCR2011 / SU47;
RX MEDLINE=21172875; PubMed=11274118;
RA Lynch D., O'Brien J., Welch T., Clarke P., Cuiv P.O., Crosa J.H.,
RA O'Connell M.;
RT "Genetic organization of the region encoding regulation, biosynthesis,
RT and transport of rhizobactin 1021, a siderophore produced by
RT Sinorhizobium meliloti";
RL J. Bacteriol. 183:2576-2585 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=211396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSyma megaplasmid";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888 (2001).
CC -!- FUNCTION: RECEPTOR FOR THE SIDEROPHORE RHIZOBACTIN.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF110737; AAD09419.1; -;
CC EMBL; AE007312; AAK65923.1; -;
CC PIR; A95420; A95420.
CC PIR; T46821; T46821.
CC InterPro; IPR000531; TonB_boxC.
CC Pfam; PF00593; TonB_dep_Rec; 1.
CC PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.

DR PROSITE; PS01156; TONB DEPENDENT REC 2; FALSE NEG.
KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
KW Plasmid; Complete proteome.
PT SIGNAL 1 26 POTENTIAL.
PT CHAIN 27 746 RHIZOACTIN RECEPTOR.
PT SITE 40 47 TONB BOX.
PT SITE 729 746 TONB C-TERMINAL BOX.
SQ SEQUENCE 746 AA; 80633 MW; 16AE44A4025D5B23 CRC64;

Query Match 3.9%; Score 206; DB 1; Length 746;
Best local similarity 20.4%; Pred. No. 0.00021;
Matches 139; Conservative 96; Mismatches 253; Indels 194; Gaps 30;

QY 101 LEQVVLGYGKGKSTVSGSAVAKVSEK--AERPVANIMDALGVAGQVMTSGDPT 159
DB 39 LEEIVVTGGSAAQISIAITIIYVDSQIQAEARSKTIQIILGET-----IPSFDA 92
QY 160 AVASVEIHGTGSLGASSAPLYVDG-----MOTSLDVVATVMPNDFMSYKLDASA 211
DB 93 SDGARTSFGQNLRG--RPPLILVDGSMNSARSLSRQFDAL---DPFNIERVEVL--SGA 145
QY 212 TSIYGARAANGVFIQTKGKMSERGRITFNASVIGISQILNTRKPLDNMTGDELLDFQVK 271
DB 146 TAIYGGNATGGINIITKGGKDAEPG-----LHAEVTTGG-----MG 181
QY 272 AGFGNNQTVQKVKMILAGAE-----LYGN--YDSLKDEYGKTLFP-----VDENHD 318
DB 182 SGFAGSQDFRNAAGAVTVNSNDARLSAGNTGATGAGYDQSGVLLIPDITQTSTAFNER 241
QY 319 ADWLKALFKTAPTSQGDISFSG-----GSQGTSYVASYGYFDQGMAREPANFKRYSGLN 374
DB 242 IDLMSIGYQIDDDR--RVFESGQYFDKQSDYGLYGGP--PAALADPSLPETRS-- 295
QY 375 PESRINWLKVGANLSCAIANRRA-----DYFGKYMGSTGTFGLVTPRYNPDV 426
DB 296 YESDFNP-----QTRSMNLNVTYDNDVFCQQLLQGSY--RTERIKFHPFPA 341
QY 427 NGDLADVVYMGATRPSTPEYPAKMPFSESHEQANVNGPAQITPIKGLTLKAQAGVDI 486
DB 342 SGNETGPFYVGS---SQDQDYGIIRALVAEPDA-----LKITYGIDA 383
QY 487 TINTTSKRPNPNPYDSTPLGRERRARYDSK-----SPTNTAAYKFSIDKHD 536
DB 384 DMSDFTARQ---NIFDMVAAGSQGLDPTNTIGTKGLYPSIDVSTVAGFAASYEATDRLT 440
QY 537 LTALMGHEYIYE--GDVIGASSKGFESDKMLLSQKTKGNSLPEHRVAEYVLSFSPR 595
DB 441 LMGVRYQFVNTEVSDPFGAAQ-----VAIIQGRATSADTTPG----- 479
QY 596 FNYGFDKMWYIDFSVRNDQSRFGSNRSWAFYVSGGMFDIYNKFIQESNWLSDLRL--- 652
DB 480 -----GEVYDIALFAGATYQILNTQVYANFSQGFELPDP 516
QY 653 KMSYGTGNSBIGNY---HQAIVTVNNYTTEDAMGLSISTAGNPDLISWKSQSFNFG-- 706
DB 517 AKYVG-----IGNYSFGGHYTLVNSV-----VGSALAEIKTNSFEIGYR 558
QY 707 LAAGAFNNLSAEVDFVVRTN 728
DB 559 LDDGTFNLETAAYYSLSDRSIN 580

RESULT 2
CIRAC ECOLI STANDARD; PRT; 663 AA.
AC CIRAC ECOLI
AC P17315;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Colicin I receptor precursor.
GN CIRAC OR CIR OR FEUA OR B2155.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89123100; PubMed=2644220;
RA Nau C.D., Konisky J.;
RT "Evolutionary relationships between the TonB-dependent outer membrane
RT transport proteins: nucleotide and amino acid sequences of the
RT Escherichia coli colicin I receptor gene.";
RL J. Bacteriol. 171:1041-1047(1989).
RN [2]
RP REVISIONS.
RA Nau C.D., Konisky J.;
RL J. Bacteriol. 171:4530-4530(1989).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / BHD2600;
RA Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,
RA Church G.M.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MGI655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glaesner J.D., Rode C.K., Mayhew G.P.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RL "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [5]
RP SEQUENCE OF 1-59 FROM N.A., AND SEQUENCE OF 26-46.
RX MEDLINE=88058737; PubMed=3316180;
RA Griggs D.W., Tharp B.B., Konisky J.;
RT "Cloning and promoter identification of the iron-regulated cir gene
RT of Escherichia coli.";
RL J. Bacteriol. 169:5343-5352(1987).
RN [6]
RP SEQUENCE OF 1-125 FROM N.A.
RC STRAIN=BPR2;
RX MEDLINE=92250419; PubMed=1315732;
RA Steffes C., Ellis J., Wu J., Rosen B.P.;
RT "The lypS gene encodes the lysine-specific permease.";
RL J. Bacteriol. 174:3242-3249(1992).
RN [7]
RP SEQUENCE OF 1-21 FROM N.A.
RX MEDLINE=90264362; PubMed=2160948;
RA Griggs D.W., Kafka K., Nau C.D., Konisky J.;
RT "Activation of expression of the Escherichia coli cir gene by an
RT iron-independent regulatory mechanism involving cyclic AMP-cyclic
RT AMP receptor protein complex.";
RL J. Bacteriol. 172:3529-3533(1990).
CC -I- FUNCTION: NOT YET KNOWN. POSTULATED TO PARTICIPATE IN IRON
CC TRANSPORT. OUTER MEMBRANE RECEPTOR FOR COLICINS IA AND IB.
CC -I- SUBCELLULAR LOCATION: Outer membrane.
CC -I- INDUCTION: BY IRON AND BY A CYCLIC AMP/CYCLIC AMP RECEPTOR
CC PROTEIN COMPLEX.
CC -I- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/ebis-sib.ch/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; J04229; AAA75183.1; -;
CC EMBL; U00007; AAA60531.1; -;
CC EMBL; AE000304; AAC75216.1; -;
CC EMBL; M19295; AAA23581.1; -;
CC EMBL; M89774; AAA17054.1; -;
CC PIR; B64984; QREIC.

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)

CC -----
DR EMBL; M31227; AAA26380.1; --
DR PIR; A41477; A41477; Autotransport.
DR InterPro; IPR006315; Autotransporter.
DR Pfam; PF03797; Autotransporter; 1.
DR TIGRfam; TIGR01414; autotrans_bar1; 3.
KW Antigen; Repeat; Signal; Cell wall; S-layer; Glycoprotein.
FT SIGNAL 1 28
FT CHAIN 29 2249
FT DOMAIN 212 1180
FT REPEAT 212 286
FT REPEAT 212 286
FT REPEAT 287 358
FT REPEAT 359 430
FT REPEAT 431 505
FT REPEAT 506 577
FT REPEAT 578 652
FT REPEAT 653 724
FT REPEAT 725 799
FT REPEAT 800 874
FT REPEAT 875 949
FT REPEAT 950 1021
FT REPEAT 1022 1093
FT REPEAT 1094 1165
FT REPEAT 1166 1180
SQ SEQUENCE 2249 AA; A9D6646C089DF087 CRC64;

Query Match 3.0%; Score 161.5; DB 1; Length 2249;
Best Local Similarity 19.4%; Pred. No. 0.35;
Matches 218; Conservative 152; Mismatches 444; Indels 311; Gaps 50;
QY 14 SIGMAQRNVKGVTSSEDEPLIGANVVVGN--TTGAATDLDGNFTLSVPNAKML 72
DB 516 SVG---AGTATLGGAIVKATTKLTDASAIVKFTNPVVTGAIDNTGNANGI----- 565

QY 73 RVSYSGMTTKEVAIANVMKIVLDPDSKVLEQVVVLGYGTGQKLSVSGSAKVSSEKLA 132
DB 566 -VTFTGNSTVTGIDGNTNSLA-----TISVGAGT---ATLGGAIVKATTKLTN 610
QY 133 KPVANIMDALQGVAGQVMTSGDPTAVASVEIHG-----TGSLSGASSAPLYIVDGMQT 187
DB 611 AASVLTLTNANAVLTGAIDNTTGGDNVGV--LNLGALSQVTDGIDGNTNSLATISVGAGT 668
QY 188 SL---DVVATMPNPFESMSVLKASATSIYGA-----RAANGVFIQTKKGMSEGRHI 239
DB 669 ATLGGAIVKATTKITNAVSARKFTNPVVTGAIDSTGNANGIV---TFTGNSTVTGDI 725
QY 240 -----TFNASVGISQI-----LNTKPLDNNMTGDELL 266
DB 726 GNTNALTAVNAGATATLGGAIVKATTKLTNAASVLTNANAVLTGAIDNTGSDNVG 785
QY 267 DFQVK-----AGFMGNNTQVKQKMDILAGAEELYG-----NYDSLKDRY 306
DB 786 VLNLGALSQVTDGIDGNTNSLATIS--VGAGTATLGGAIVKATTKLTNAASVLTITNAN 843
QY 307 GKTLFPVDNHDADWLKALFKTAPTSQ--GDISFGSGSGSOTSYASIGYFDGEGMAR-EP 363
DB 844 AVLTVGAVDNTTGGDNVGVNLGALSQVTDGI-----GNTNSLATISVG-----AGTATLGG 895
QY 364 ANFKRYSGRINPESRINELKVGANLSGAIANRRSADYEGKYMGSGTGVLTMPRYNP 423
DB 896 AVIKATTKLTNAASVLTITNANAVLTGAIDNTGDD-----NVGVLT----- 937
QY 424 FDVNGDLADVYMYGATRPSTMPYPAKMPFSSSHQANVNGFAQITPIKGLTLKAQAG 483
DB 938 -NLGALSQVTDGIDG-----NTNSLATISVGAGTATLGGAIV 972
QY 484 VDIITWR---TSSKMPNPNPDSTPLGERREARVDRVSKFTNVAEYKFSIDEKHDL-TA 539
DB 973 IKATTKLTDAASAVKFTNPVVTGAIDNTGNANGI--VTFTGNSTVTGNTGNATATV 1031

QY 540 LMGHEYIIEYEGDVI-----GASSKGFESDKMLLSQKGTGNSLSLPEHRVAEYAYL 590
DB 1032 NVGAGLLQVQGVVKKANTINLTNDAVTFNTPVVVTGAIDNTGNA----- 1077
QY 591 SPFSRFNYGFDKWMYIDFSVRNDQSRSGNSRNSAFYSVGGMFDIYNKFIQESNW-LSD 649
DB 1078 -----NNGI-----VTFTGNSTVTGNTGNATATVNVGALLQVQGVVKKANTINLTN 1126
QY 650 LRLKMSYGT-----TGNS-----EIGNYNHQAALTVNNYTEDAMG 684
DB 1127 NASAVTFNTPVVVTGAIDNTGNANGIVTFTGNSTVTGIDGNTN--ALATVN-----VG 1178
QY 695 LSIIS-----TAGNPDLSEKQSQNFGLAAGAFNNRLSARVDYF---VTTTNDMLID 733
DB 1179 AGITLQAGSGLAANNIDFGARSTLEFNGPLDGG-----GKAIPIYFKGAIANGNAILN 1232
QY 734 VPMPYISGFFSQYQNVGSMKNTGVDSLKGTIYQN-KDNNVYASANFNFNROEITKLP--- 790
DB 1233 VNTKLT---ASHLTIGTVAEINIGAGNLFTIDASVGDVTTILNAQINFRARDSVLVLSN 1289
QY 791 ---FGLNKYML-----PNTGTIWEIGYPSFYM-AEYAGIDKKTGQLWVVPQVDA 838
DB 1290 LTGVGVNNILLAADLVAPGADGIVFNGGVNGLNVGNAVGTARNIG-----DG 1339
QY 839 DGNKVTTSQYSADLETRIDKSVTPPTIGGFSI-GASWKGSLSDADP---AYIVGKWMIN 893
DB 1340 GGNKENT-----LLIYNVAVT---ITDDVNLGEGIQNVLINKNADFTSTAFNAGAIQIN 1389
QY 894 NDRIYTE-----NAGGLMQLNKDKMLNNAWTEENKEDTVPKLGQSPQFTHLLENAS 945
DB 1390 DATYTTIDANNGLNIPAGNIQFAHADAAQLVQLNSSGNDRTITLGNIDPDNDEGI---- 1445
QY 946 FLRLKQLKLTYYLNSLPAQNV-IGGARVYLMARNALTVTKYKG 989
DB 1446 -----VILNSVTAGKLTAGKTFGGAHKLQTL-FKG 1478

RESULT 11
OAR_MYXXA
ID OAR_MYXXA STANDARD; PRT; 1061 AA.
AC P38370;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE OAR protein precursor.
GN OAR.
OS Myxococcus xanthus.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Cystobacterineae; Myxococcaceae; Myxococcus.
OX NCBI_TaxID=34;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 776-796.
RC STRAIN=DZF1;
RX MEDLINE=93328680; PubMed=8335633;
RA Martinez-Canamero M., Munoz-Dorado J., Farez-Vidal E., Inouye M.,
RA Inouye S.
RT "Oar, a 115-kilodalton membrane protein required for development of
RT Myxococcus xanthus.";
RL J. Bacteriol. 175:4756-4763(1993).
CC -!- FUNCTION: REQUIRED FOR CELLULAR ADHESION DURING FRUITING BODY
CC FORMATION.
CC -!- SUBCELLULAR LOCATION: OUTER-MEMBRANE ASSOCIATED.
CC -!- SIMILARITY: LOCAL, TO TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)

QY 898 FTE-----NAGGLMQLNKDKMLNNAWTEDEKEDTDPVKLGOSQPDTHLEENASFLRL 949
 Db 1166 TIDANNGLNIPAGNIKFAHAQALILQNSGNDRTITLGNIDPDNDDEGI----- 1217
 QY 950 KNLKLTLYLPSLPAGQNV-IGGARVYLMARNLTVTKYKG 989
 Db 1218 -----VILNSVTAGKLTAGGKTGGAHKLQDIV-FKG 1250

RESULT 13
 ID IRGA VIBCH STANDARD; PRT; 652 AA.
 AC P2772; O9KUP0;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Iron-regulated outer membrane virulence protein precursor.
 GN IRGA OR VC0475.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Classical Ogawa 395 / ATCC 39541 / Serotype O1;
 RX MEDLINE=93023868; PubMed=1406279;
 RA Goldberg M.B., Boyko S.A., Butterton J.R., Steebner J.A.,
 RA Payne S.M., Calderwood S.B.;
 RT "Characterization of a Vibrio cholerae virulence factor homologous to
 RT the family of TonB-dependent proteins.";
 RL Mol. Microbiol. 6:2407-2418(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=El Tor N16961 / Serotype O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolova M.D., Vamathevan J., Bass S., Qin H., Dragoi L., Sellers P.,
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 RT cholerae.";
 RL Nature 406:477-483(2000).
 RN [3]
 RP SEQUENCE OF 1-151 FROM N.A.
 RC STRAIN=Classical Ogawa 395 / ATCC 39541 / Serotype O1;
 RX MEDLINE=91072235; PubMed=2174861;
 RA Goldberg M.B., Boyko S.A., Calderwood S.B.;
 RT "Transcriptional regulation by iron of a Vibrio cholerae virulence
 RT gene and homology of the gene to the Escherichia coli fur system.";
 RL J. Bacteriol. 172:6863-6870(1990).
 CC -1- FUNCTION: INVOLVED IN THE INITIAL STEP OF IRON UPTAKE BY BINDING
 CC FERRIC VIBRIOBACTIN, AN IRON CHELATING SIDEROPHORE THAT ALLOWS
 CC V. CHOLERA TO EXTRACT IRON FROM THE ENVIRONMENT.
 CC -1- SUBCELLULAR LOCATION: Outer membrane.
 CC -1- MISCELLANEOUS: REGULATION OF THE IRGA EXPRESSION IS NEGATIVELY
 CC REGULATED AT THE TRANSCRIPTIONAL LEVEL BY IRON.
 CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U72152; AAC4766.1; -;
 CC EMBL; AE004134; AAF93648.1; -;
 CC PIR; D82317; D82317.

DR PIR; S25265; S25265.
 DR HSSP; P05825; 1FEP.
 DR TIGR; VC0475; -;
 DR InterPro: IPR000531; TonB_boxC.
 DR Pfam: PF00593; TonB_dep_Rec; 1.
 DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; 1.
 DR PROSITE: PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Virulence; Outer membrane; Iron transport; Transport; TonB box;
 KW Signal; Receptor; Complete proteome.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 652 IRON-REGULATED OUTER MEMBRANE VIRULENCE
 FT SITE 33 40 TONB BOX.
 FT SITE 635 652 TONB C-TERMINAL BOX.
 FT CONFLICT 294 294 D -> G (IN REF. 1).
 FT CONFLICT 448 448 K -> Q (IN REF. 1).
 FT CONFLICT 502 502 A -> T (IN REF. 1).
 SQ SEQUENCE 652 AA; 71669 MW; A95F82FEC072EC93 CRC64;
 Query Match 3.0%; Score 159; DB 1; Length 652;
 Best Local Similarity 19.6%; Pred. No. 0.086;
 Matches 144; Conservative 86; Mismatches 248; Indels 258; Gaps 34;
 QY 98 SKVLEQVVVLGYGTGQKLSVSGSAKVSSEKLAEPKPAVIMDALQGVAGMQVMTTSGD 157
 Db 29 TKTDETMVVTAGYAQVIQNPASISVISREDLESRYRDVTDALK-SVPGVTV-TGGGD 86
 QY 158 PTAVASVEIHGTGSLGASSAPLYVDG-MQTSLDVVVATWPNND-----FE 201
 Db 87 TT----DISRGMG-----SNYTLILVDGKRQTSRQT----RPNSDGPGEQWLPPIQAI 135
 QY 202 SMSVLKDSATSIVGARAANGVVFIQTKGKMSERGITF-----NASYGISQILN--- 252
 Db 136 RIEVIRGPMST-LYGSDAIGGVINIITRKDQOQWSGNVQLSTVVOENRSGDSQANFFV 194
 QY 253 TKPLDNNMT-----GDELL-----DFQVKAGFGWNN 278
 Db 195 TGPLSDALSILQVYQTTQDEDEIEHGYGDKLSRLTSKLYNLQNPDLHQLEAGVSAQD 254
 QY 279 QTVQVKMDILAGAEDLYGNYSLKDEYKTLFPVDFNHDADWL----- 322
 Db 255 RENNKGSAQSGCRGTCSTND---NOYRNR--HVAVSHQGDMDQVQSDTYLQYEENTN 309
 QY 323 -----KALPKT---APTSQDISFSGSQG-----TSYAS-----IGY 353
 Db 310 KSREMSIDNTVFASTLVAPIGEHMLSF--GVEKGHESLEDKTSNKISSRTHISNTQWAGF 367
 QY 354 FDOEGMARBPANPKRYSGRLNPFESRINELKVGANLSGAIANRRRSADYFGKYMGSGTGF 413
 Db 368 IEDEWALAE-----QFRLTFGGRLDHDKNYSGHFS----- 397
 QY 414 VLTMPRYNPPFDVNGDLADVYMYGATRPSMTPEYFAKMRP-FSSSESHQANVNGPQAITP 472
 Db 398 -----PRVYGVW-----NLDPILMTVKGGVSTGFAPQLREVTPDMQGVSGGGNIYGNPDLKP 449
 QY 473 -----IKGLTLKAQAGVDITNRTSSKR-----MNNPYDSTPLGERRERAYR--- 515
 Db 450 ETSINKELSLMYSTGSLAASLTAFHNDPKDKITRVACPAICITAGPNQWGAAPTTRVNI 509
 QY 516 DVSKSTNTAEYKFSIDEXHDLTALMGHEVIEYEGDVIGASSKGFESDKMLLSQGKTGN 575
 Db 510 DEATYGAETLSLPITESVELSSVYTH-----SEKSGSN 546
 QY 576 -----SLSLPEHRAVEAYLSPFRFNYGDKMYIDFSVRNDQSSRFSGNN-----RSAW 626
 Db 547 FAGRLPILQPKHLFNANLSWQTTDRLN-----SWANLYRGKEMQPEGGASNDFFIAPSYT 602
 QY 627 FYSVGGMF-----DIYNKFIQESNWLSDLRLKMSYGTGTGNGEIGNYNNHQAALVTVN 676
 Db 603 FIDGTGYALTDTATTKAAVYNLFDQEVNY-----AEYV----- 636
 QY 677 NYTDEA-----MGLSIS 688
 ||| :|||

```

Db 637 -YVEDGRRYWLGLDIA 651
RESULT 14
TOXA CLODI
ID TOXA CLODI STANDARD; PRT; 2710 AA.
AC P16154;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Toxin A.
GN TOXA OR TCDA.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RX MEDLINE=90221894; PubMed=2109310;
RA Sauerborn M., von Eichel-Streiber C.;
RT "Nucleotide sequence of Clostridium difficile toxin A.";
RL Nucleic Acids Res. 18:1629-1630(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RX MEDLINE=90129305; PubMed=2105276;
RA Dove C.H., Wang S.Z., Price S.B., Phelps C.J., Lyster D.M.,
RA Wilkins T.W., Johnson J.L.;
RT "Molecular characterization of the Clostridium difficile toxin A
RT gene.";
RL Infect. Immun. 58:480-488(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RA von Eichel-Streiber C.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -! DOMAIN: THE C-TERMINAL PART OF TOXIN A CONSISTS OF A 833 AA
CC REPETITIVE STRUCTURE. THIS PART OF TOXIN A IS COMPOSED OF FIVE
CC DIFFERENT OLIGOPEPTIDES.
CC -! DISEASE: CLOSTRIDIUM DIFFICILE PRODUCES TWO TOXINS, AN
CC ENTEROTOXIN CALLED A AND CYTOTOXIN B. ONLY AFTER THE ENTERAL
CC DELIVERY OF THE ENTEROTOXIN A MAY THE CHARACTERISTIC DISEASE
CC CALLED PSEUDOMEMBRANOUS COLITIS BE INDUCED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X51797; CAA36094.1; -
DR EMBL; M30307; AAA23283.1; -
DR EMBL; X92982; CAA63564.1; -
DR InterPro; IPR002479; CW_binding.
DR Pfam; PF01473; CW_binding_1; 28.
DR Pfam; PF04488; Gly_transf_sug; 1.
KW Toxin; Enterotoxin.
SQ SEQUENCE 2710 AA; 308052 MW; 0A6852CB84C14421 CRC64;

Query Match
Best Local Similarity 3.0%; Score 158.5; DB 1; Length 2710;
Matches 251; Conservative 166; Mismatches 429; Indels 441; Gaps 74;

QY 21 QNRTV-KGTVISSENEPLIGANVVVGNITIGATD-----LDGNFTLSVP 66
Db 1359 ENGTTKKGKLIKDVLSKIDINKKLIIGNQTFDSDIDNKRDIYFLCELDKLSLIE 1418
QY 67 AN-AKMLRVSGMTTKEVAIANVWKIVLDPDSKVLBOVVLYG-----TGQKLS 116
Db 1419 INLVAKSYSLLSG--DKNYLISNL-----SNTIEKINTGLDLSKNIAVNYTDESN 1468

```

```

QY 117 TVSGSAKVSSEKLA--EKPVANIMDALQ-----GOVAGMQ 150
Db 1469 KYFGAISKTSQKSIHYKDKSKNILEFYNDSTLENSKDFIAEDINVPKMDINITGKY 1528
QY 151 VMTSGDPTAVASVEIHGTSGLSGASSAPLIYVDGQTS-LDVVATM-----NPNDFESMV 205
Db 1529 YVDNNTDKSIDFISLSVSKNQVKNG--LYLNESVYSSYLDPVKNSDGHHTSNFWNL-F 1585
QY 206 LKDSATSICYARAANGV--FIOTKKGKMSERGITF-----NASYCI 247
Db 1586 LDNISFWLFGFENINVIDKYF--TLVGK-TNLGYVEFICDNKNIDYIFGEWKTSSK 1642
QY 248 SQILN-----TKPLDNMTGDEL---LDFOVKAGFGNNGQTVOKVXDMILAGAEGLVG 297
Db 1643 STIFSGNGRVVVEPIYNPDGEDI STLDSYEP-LYGDRIYINKV-LI--APDLYT 1696
QY 298 NYDSLK-DEYKGTLP-----VFNHDAWMLKALFKTAPTSGD-----335
Db 1697 SLININTNYSNEYYPEIIVLNPNTFHKKNINLDS-----8SFYKWSGSDGDFILVRYL 1752
QY 336 -----ISPSGSGSGTSY-----ASIGY-----PDQEG-MAREPAN 365
Db 1753 RESNKKILQIRIKGILSNTOSFNQMSIDFKDIKKLSLGYIMSNPKFSENELDRDHUG 1812
QY 366 PKRYSGRLNFESRINELWLVKANLSGAIANRRSADYFG-----KYMGS 409
Db 1813 FKIIDNKYYVDEDSKLVK-----GLNINNSLFYDPDIEPNLVGTQTINGKYYFDI 1866
QY 410 GTFGVLTPRYYN--PFDVNGDLADVYMYGATPSMTEPYPAKMRPFSSSHOANVNGF 467
Db 1867 NTGAALTYSKIINGKHFFNND--GVMQLGVFKGPDGFE-YFA---PANTQNN--NIEQ 1918
QY 468 AQITPIKGLTL-----KAQAGVDITNTR-----TSKRMPPNPYDST 504
Db 1919 AIYQSKFTLLNGKKYFDNNSKAVTGWRIINNEKYFNPNNIAAVALQVIDNNKYYFN 1978
QY 505 PLGERREYARDVSKSF--TNTAEYKFSIDKHDLTALMGHEYIEYEGDVI GASSKGPS 562
Db 1979 P-----DTAI--ISKWQTVNGSRYYFTD-----TAFNGYKTDG-----KHFYDS 2021
QY 563 DKMLLSQGTGNSLSLPEHVAEY-----AYLSFFSRFN--YGFDK- 602
Db 2022 DCVVKIGVFSTNGF-----EYFAPANTVNNIEGOAIVYQSKFTLLNGKKYFDNN 2073
QY 603 -----WNYID-----FSVRNDQSSR-----FG 619
Db 2074 SKAVTGLQTDISKYYFNTNTAEATGWTIDGKKYFNTNTAEATGWTIDGKKYFN 2133
QY 620 SN-----NRSAMFYSGVMFDI-----YNKFIQESNWLSDLRLKMSYGTGNSE 663
Db 2134 TMTAISTGYTIINGKHFFINTDGMQIGVFKGPNGF-----EYFAPANTD 2179
QY 664 IGYNHQALVTNNY--TEDAMGLSISTAGNPDLSW-----EKQSQFNFLGACAFNNRLSA 718
Db 2180 ANNIEGOAILYQNEFLTLLNGKKYFGSDSKAVTGWRIINNNKYYFN-----PNNATA 2232
QY 719 EVDFTVRTNDMLIDVPMPIYSGFPSQYQNVGSMKNTGVDLSLKGTTIYQNKDNNVYASAN 778
Db 2233 ---IHLCTINN-----DKYFYSYD--GILQNGYI-----TIBRN--NFYFDAN 2268
QY 779 FNYNRQETIKLFFGLN--KYMLP-----NTGTIWEIGYPNSFYMAEYAGIDKTKGKQLWY 831
Db 2269 -NESQ--VTGVFKGPNGFEPAPANTHNNNIEGOAIVYQNKFLT-----NKKKYF 2318
QY 832 -----VPGVDADGNKVTTSQYSADLETRIDKSVTPPITGG-----WSLGS 877
Db 2319 DNDKAVTGMQTDGKK-----YFNLNTAEATGWTIDGKKYFNLNTAEATGWO-- 2371
QY 878 SLDDADFAIVIGKWMIN-----NDRYFTEYENAGGLMQL-----NKOKMLNMTEDNKE 924
Db 2372 TIDGKKYFNTNTFIATGYTSINGKHFFINTDGMQIGVFKGPNGFEPAPANTDANNI 2431

```

QY 925 TDVFKLQSQPQFDTHLENAFLRLKLNKLTYYLPNSLFAQONVIGGARVYLMARNLLTV 984
 Db 2432 E-----GQA-----ILYQNKFLTLNGKYYFGSDSKAVTGLRTIDGKYYFNTTAVAV 2480
 QY 985 TKYKGFPEAGNVGNKQYNSQYVA 1011
 Db 2481 TGMQTIN-----GKKYFNTNTSIA 2500

RESULT 15

Y262_HAEIN STANDARD; PRT; 723 AA.
 AC P44600;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable tonB-dependent receptor HI0262 precursor.
 GN HI0262.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Karlavage A.R., Bult C.J., Tomb J.-P., Dougherty B.A., Merrick J.N.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geochagen N.S.M.,
 RA Gnehm C.B., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RT Rd.";
 RL Science 269:496-512(1995).
 RN [2]
 RP IDENTIFICATION BY MASS SPECTROMETRY.
 RX MEDLINE=1037488; PubMed=10675023;
 RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
 RA Gray C., Fountoulakis M.;
 RT "Two-dimensional map of the proteome of Haemophilus influenzae.";
 RL Electrophoresis 21:411-429(2000).
 CC -1- FUNCTION: PROBABLE RECEPTOR, TONB-DEPENDENT.
 CC -1- SUBCELLULAR LOCATION: Outer membrane (potential).
 CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>/
 CC or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL; U32712; AAC21927.1; -;
 DR PIR; C64058; C64058.
 DR TIGR; HI0262;
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_dep_Rc; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Outer membrane; Receptor; Signal; TonB box; Complete proteome.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 723 PROBABLE TONB-DEPENDENT RECEPTOR HI0262.
 FT SITE 706 723 TONB C-TERMINAL BOX.
 SQ SEQUENCE 723 AA; 80775 MW; 91EB3AB0FFEA2984 CRC64;

Query Match 3.0%; Score 157; DB 1; Length 723;
 Best Local Similarity 17.5%; Pred. No. 0.13;
 Matches 151; Conservative 125; Mismatches 269; Indels 316; Gaps 38;

QY 124 KVSSEKLAEPKPVANIMDALQGVAGQVMTSGDPTAVASVEIHGTGSLGASSAP----- 178
 Db 46 KQSKDLSLSKQATSVADALE-----IPNDVRG-GSRSIQAKPNIRGL 88
 QY 179 -----LYIVDGQMTSLDVVATNPDPFESMVLKD-----ASATSIYGARAAANGVVFIQT 228
 Db 89 SDNRVVQVIDGVRQNFDL--AHRGSYFLPMSLIQEIVIKGPPSSSLWGSALGVVAMRT 146
 QY 229 KKG-----KMSERGRITFNASYGISQILNTKPLDNNMTGDELLDFQVKGAGWGNQTVQKV 284
 Db 147 PNALDLLKNDKFGVKIROGYQTANNLSKDVSVFAANDKF----- 187
 QY 285 KOMILAGAEDLYGNYDSLKDEYKTLFPVDFHHDADWLKALPKTAPTSGDISFSGSGSQ 344
 Db 188 -DVLISG---FYNNADNLRTGKGNL-----NNTA----- 213
 QY 345 TSYIASIGYFDQEGMARBPANFKRYSGRLNFSRINEWLKVGANLSGAJANRRSADYFGK 404
 Db 214 ---YKQFGGLAKFGWQINDANRVELSHR---ETRFKQTAPSNNEVENELTNEQITDQIKK 267
 QY 405 YMGSGTFFGLTMPRYNPFVNGDLADVYMYGATPMSMT---PYFAKWRPFSSSHQA 462
 Db 268 FH-----GQKDDL--LPPTTQPSERSESEFYKVK----- 295
 QY 463 NVNGFQAITPIKGLTLKAQAGVDITNTRTSSKMPNPNPYDSTPLG-----ERRR 512
 Db 296 -----TRLSGSVLTQQIPDQSTVFNLYLTDPNLYNLTHIALYNNKTIEQQRKVS 347
 QY 513 AYRDVSKSTNTAAYKFSIDEXHDLPALMGHEY-----IEYEGDVIQASSKGFSDKMLLL 568
 Db 348 GVKDQTKLTTRGINLNRSSLSLH-ISFVYGVDMRDKIRTERGTGNSDAK-FRADPVN-A 404
 QY 569 SQKTKG-----NSLSLPEHRAVEY-----AYLSFFSRFNYGFDKWMY 605
 Db 405 NSNTTGVYLIHPIPLFGEKLLVSPSYRHYDTSSTKTYKDNHLSPATKLTIVTNW-- 462
 QY 606 IDFSVRNDOSR-----FGSNRSAMFYSGVMFDIYNKFIQESNWLSDRL 652
 Db 463 LDFTAKYNEAFRAPSQMERFVSGAHFGAN-----TLG--LDHINRFVANPN---LRP 509
 QY 653 KMSYGTGTGNSSEI-GNYNHOAL-----VTVNNYTEDAMGLSISITAGNPDLSWEKQ 700
 Db 510 E-----TAKNKEITANLHFDLSLFGQDKFKIEATYFRNDVKDFINLKIFNDAKTSAS-- 561
 QY 701 SQPNFGLAAGAFNNRLSAEVDYFVRTNDMLIDVPMPIYSGFESQYQVNGSMKNTGVLS 760
 Db 562 -----AGANPN-----TNGALLPN-----SOYQNIITNARLSGIELQ 593
 QY 761 LKGTIYQNKDMNVYASANFYNNRQETKLPFGLNKYLMPNTGTIWEIGVPSNFYMAEYAG 820
 Db 594 AQ---YQTERLTULTN-----YGSTK-----G 612
 QY 821 IDKTKGQQLWYVPGQVDADGNK-----VTTTQYSDADLETIDKSVTTPITGGFS 869
 Db 613 KDKSGEALSNIAASKIGGVNVALVKDKFTVGTATVTHAAQRRVPKDHSTVTP----- 666
 QY 870 LGASWKGSLDADFAIVGKW 890
 Db 667 ---SYILTDLRATYAPLKGW 684

Search completed: January 7, 2004, 19:03:56
 Job time : 38 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2004, 18:50:04 ; Search time 51 Seconds
(without alignments)
5145.873 Million cell updates/sec

Title: US-09-581-286A-424
Perfect score: 5298
Sequence: 1 MKRWTLFFCLLTSTIGWAMA.....VGKNQYPNSQYVAGIQLSF 1017

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPREMBL_23.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_prodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5286	99.8	1017	Q9ZA60	Q9za60 porphyromon
2	1001	18.9	1137	Q8VTA9	Q8vta9 bacteroides
3	944	17.8	1038	Q45780	Q45780 bacteroides
4	904	17.1	1089	Q93PC3	Q93pc3 microscilla
5	843	15.9	947	Q93FV1	Q93fv1 bacteroides
6	734.5	13.9	1125	Q93TH9	Q93th9 bacteroides
7	645.5	12.2	1089	Q9AF86	Q9af86 bacteroides
8	302	5.7	59	Q9XAU9	Q9xau9 porphyromon
9	270	5.1	844	Q8KB41	Q8kb41 chlorobium
10	232.5	4.4	757	Q9KIB0	Q9kib0 porphyromon
11	226.5	4.3	616	Q91473	Q91473 pseudomonas
12	225.5	4.3	867	Q9KIB1	Q9kib1 porphyromon
13	222	4.2	1022	Q8PI21	Q8pi21 xanthomonas
14	210	4.0	966	Q8P853	Q8p853 xanthomonas
15	207	3.9	696	Q9PPG3	Q9ppg3 campylobact
16	201	3.8	2057	Q9RE05	Q9re05 leuconostoc

17	200.5	3.8	687	2	Q9XDI8	Q9xd18 bacteroides
18	200.5	3.8	812	16	Q8XSQ1	Q8xsq1 ralstonia s
19	199	3.8	696	2	O07651	O07651 campylobact
20	197.5	3.7	713	16	Q9KNA1	Q9kna1 vibrio chol
21	196.5	3.7	1307	16	Q8EV70	Q8ev70 mycoplasma
22	196.5	3.7	2143	16	Q8RIF5	Q8rips fusobacteri
23	196	3.7	731	16	Q8D192	Q8d192 yersinia pe
24	196	3.7	735	2	Q45341	Q45341 bordetella
25	195.5	3.7	715	16	Q8RFP8	Q8rfd8 fusobacteri
26	195	3.7	754	2	P70847	P70847 bordetella
27	195	3.7	991	16	Q8F194	Q8fi94 leptospira
28	194.5	3.7	682	2	Q93FB0	Q93fb0 escherichia
29	194.5	3.7	779	16	Q8KFA4	Q8kfa4 chlorobium
30	194.5	3.7	833	2	Q9KIB2	Q9kib2 porphyromon
31	194	3.7	964	16	Q8PJU0	Q8pjk0 xanthomonas
32	194	3.7	3692	2	Q8KRR3	Q8krr3 fusobacteri
33	192.5	3.6	682	2	Q8VR48	Q8vr48 escherichia
34	192.5	3.6	682	16	Q8FAQ7	Q8faq7 escherichia
35	192	3.6	659	16	Q8XG33	Q8xg33 escherichia
36	192	3.6	1536	2	Q48031	Q48031 haemophilus
37	191.5	3.6	718	16	Q8D5A8	Q8dsa8 vibrio vuln
38	191	3.6	690	16	Q8CVW8	Q8cvw8 escherichia
39	190.5	3.6	963	16	Q8P5M2	Q8p5m2 xanthomonas
40	190	3.6	599	16	Q8YH18	Q8ynh8 brucella me
41	189.5	3.6	1758	16	Q9JMS5	Q9jms5 escherichia
42	189	3.6	708	2	Q9XBV1	Q9xbv1 porphyromon
43	187.5	3.5	657	16	Q8RHN9	Q8rhn9 fusobacteri
44	187	3.5	3988	17	Q8TPZ1	Q8tpz1 methanosarc
45	186.5	3.5	973	16	Q8PI48	Q8pi48 xanthomonas

ALIGNMENTS

RESULT 1

Q9ZA60 PRELIMINARY; PRT; 1017 AA.
AC Q9ZA60;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Receptor antigen (Raga).
GN RAGA.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99150246; PubMed=10024556;
RA Hanley S.A., Aduase-Opoku J., Curtis M.A.;
RT "A 55-kilodalton immunodominant antigen of Porphyromonas gingivalis
RL W50 has arisen via horizontal gene transfer.";
RL Infect. Immun. 67:1157-1174 (1999).
CC -|- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR EMBL; AJ110872; CAA10226.
DR InterPro; IPR000531; TonB_BoxC.
DR Pfam; PF00593; TonB_dep_Rec; I.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
KW Membrane; Outer membrane; Receptor; TonB box.
SQ SEQUENCE 1017 AA; 112332 MW; 1B0297615DF15239 CRC64;

Query Match 99.8%; Score 5286; DB 2; Length 1017;
Best Local Similarity 99.8%; Pred. No. 1.2e-269;
Matches 1015; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	1	MCRWTLFFCLLTSTIGWMAQNRTVKGTVISSEDENEPIGANVVVGVNTTGAATDLGN 60
Db	1	MCRWTLFFCLLTSTIGWMAQNRTVKGTVISSEDENEPIGANVVVGVNTTGAATDLGN 60
QY	61	FTLSVPANAKMLRVSYSGMTTKEVAIANVMKIVLPDPSKVLEQVVVLGVTGQKLTSTVSG 120
Db	61	FTLSVPANAKMLRVSYSGMTTKEVAIANVMKIVLPDPSKVLEQVVVLGVTGQKLTSTVSG 120

QY	121	SVAKVSEKLAEPVANI	DALQOVAGMOMTTS	GDPTAVASVEIHGTGSLGASSAPLY	180
DB	121	SVAKVSEKLAEPVANI	DALQOVAGMOMTTS	GDPTAVASVEIHGTGSLGASSAPLY	180
QY	181	IVDMQTSLDVATMNP	DFESMSVLKASATS	IYGARAANGVVFIOTKKGKMSRGRIT	240
DB	181	IVDMQTSLDVATMNP	DFESMSVLKASATS	IYGARAANGVVFIOTKKGKMSRGRIT	240
QY	241	FNASYGTSQILNTKPL	DMWTGDELLDFQV	KAGFWGNNOTVQVKDMILAGAEGLYNYD	300
DB	241	FNASYGTSQILNTKPL	DMWTGDELLDFQV	KAGFWGNNOTVQVKDMILAGAEGLYNYD	300
QY	301	SLKDEYKTLFPVDFN	HDADWLKALFKTAP	TSQGDISFSGSQSTSYASIGYFDQBGMA	360
DB	301	SLKDEYKTLFPVDFN	HDADWLKALFKTAP	TSQGDISFSGSQSTSYASIGYFDQBGMA	360
QY	361	REPANPKRYGRLNFE	SRLNFWLKGANL	SGAIANRRSADYFGKYMGSGTGGVLTWPRY	420
DB	361	REPANPKRYGRLNFE	SRLNFWLKGANL	SGAIANRRSADYFGKYMGSGTGGVLTWPRY	420
QY	421	YNPPDVNGDLADVY	MTGATRPSMTPEY	FAKMRPFSSSHQANVNGFAQITPIKGLTLKA	480
DB	421	YNPPDVNGDLADVY	MTGATRPSMTPEY	FAKMRPFSSSHQANVNGFAQITPIKGLTLKA	480
QY	481	QAGVDINTRTSSKRM	PNNPYDSTPLGER	RERAYRDVSKSFTNTAEYKFSIDEKHDLTAL	540
DB	481	QAGVDINTRTSSKRM	PNNPYDSTPLGER	RERAYRDVSKSFTNTAEYKFSIDEKHDLTAL	540
QY	541	MGHEYIEYEGDVI	GASSKGFESDKML	LSQKTSLSLPEHRVAEYAYLSFSSRFNYGF	600
DB	541	MGHEYIEYEGDVI	GASSKGFESDKML	LSQKTSLSLPEHRVAEYAYLSFSSRFNYGF	600
QY	601	DKWYIDFVRNDQ	SRRFGNRSANFY	SYGCMFDIYNKFIQESNWLSDRLKMSYTTG	660
DB	601	DKWYIDFVRNDQ	SRRFGNRSANFY	SYGCMFDIYNKFIQESNWLSDRLKMSYTTG	660
QY	661	NSEIGNYNHQA	LVTVNNYTEDAM	GLSISTAGNPDLSEKOSQFNFLAAGAFNNLSAEV	720
DB	661	NSEIGNYNHQA	LVTVNNYTEDAM	GLSISTAGNPDLSEKOSQFNFLAAGAFNNLSAEV	720
QY	721	DFYVRTNDMLIDV	PMPIYISGFPSQY	QNVGSMKNTGVDLSLKGTYIQNKDMNVYASAFN	780
DB	721	DFYVRTNDMLIDV	PMPIYISGFPSQY	QNVGSMKNTGVDLSLKGTYIQNKDMNVYASAFN	780
QY	781	YNRQEIITKLPGL	NKMLPNTGTTW	EIGYPNSFWAIEYAGIDKKTGKQLWYVPGQVDADG	840
DB	781	YNRQEIITKLPGL	NKMLPNTGTTW	EIGYPNSFWAIEYAGIDKKTGKQLWYVPGQVDADG	840
QY	841	NKVTTSOYADLE	TRIDKSVTPPTTG	FSGLASWKGSLDADPAYIVGKWMNNDRYFTE	900
DB	841	NKVTTSOYADLE	TRIDKSVTPPTTG	FSGLASWKGSLDADPAYIVGKWMNNDRYFTE	900
QY	901	NAGGLMQLNKML	NLAWTENDKETD	VPKLGQSPQFDTHLENASFLRLKMLKLYVLPN	960
DB	901	NAGGLMQLNKML	NLAWTENDKETD	VPKLGQSPQFDTHLENASFLRLKMLKLYVLPN	960
QY	961	SLFAGQNVIGGAR	YVLMARNLLTV	TKYKGFDPPEAGNGVKNQYPSQYVAGIQLSF	1017
DB	961	SLFAGQNVIGGAR	YVLMARNLLTV	TKYKGFDPPEAGNGVKNQYPSQYVAGIQLSF	1017
RESULT 2					
Q8VTA9	ID	Q8VTA9	PRELIMINARY;	PRT;	1137 AA.
AC	Q8VTA9				
DT	01-MAR-2002	(TrEMBLrel. 20, Created)			
DT	01-MAR-2002	(TrEMBLrel. 20, Last sequence update)			
DT	01-MAR-2003	(TrEMBLrel. 23, Last annotation update)			
DE	FRG.				
GN	FRG.				
OS	Bacteroides fragilis.				
OC	Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;				

OC	Bacteroidaceae; Bacteroides.
OX	NCBI_TaxID=817;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=NTC 9343;
RA	Comstock L.E.;
RT	"A region of the Bacteroides fragilis chromosome containing a cluster
RL	of regulatory genes - 'fragilis regulatory region'.;"
RL	Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=NTC 9343;
RA	Coyne M.J., Comstock L.E.;
RL	Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC	-I- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR	EMBL; AF329100; AAL57000.1; -
DR	InterPro; IPR000531; TonB_boxC.
DR	Pfam; PF00593; TonB_dep_Rec; 1.
KW	Membrane; Outer membrane; Receptor; TonB box.
SQ	SEQUENCE 1137 AA; 126187 MW; E80F3FBE81D9DBF CRC64;
	Query Match 18.9%; Score 1001; DB 2; Length 1137;
	Best Local Similarity 28.5%; Pred. No. 3.3e-44;
	Matches 313; Conservative 193; Mismatches 417; Indels 176; Gaps 43;
QY	23 RTVKGTVISSEDNEPLIGANVVVGGNTTIGAAATDLDGNTLSVPANAKMLRVSYSGMTTK 82
DB	111 KVKVG-VKXDA SEALIGANVLVKGAT-GVITDMNGSELEVPGNA-VLIQITVIGYVQQ 167
QY	83 EVAIAN--VMKIVLDPDSKVLQEVVLVYGTGQKLSVSGSVAKVSEKLAEPVANIMD 140
DB	168 DVAVKRDQLAVLLKEDTKTLDVGVVVGYT-MKKXDLTGAVASVKMDDTPLSTISTVSH 226
QY	141 ALQGVAGQVMTTSGDPTAVASVEIHGTSGLGASSAPLYIVDGMQ----- 186
DB	227 ALAKGAALQVNTISAQPGGGTFRIRGAASTGAGNDPLIIVDGPVSNAGNVUSVGNSD 286
QY	187 --TSLDVATMNPDPFESMSVLKDSATSIYGARAANGVVFIOTKKGMSERGITPNAS 244
DB	287 NGTDMNILASINPDIESIEVLKDSATYI GARAGSGVLIITTKRG-EGKPKVTVSGS 345
QY	245 YGISQILNTKPLDNMTGDELL---DFQVKAGFWGNNQTVQVKQDMLAGAEULDGYNDS 301
DB	346 -----ATVQTMA TKYEMLDAQDFMIQSNRWPKWKMYDNKVGIYQGNB----- 389
QY	302 LKDEYKTLFP-----VDFNHDAWLKALFKTAPTSGDISFSGSGTSSYASIGYF 354
DB	390 --SEAGSAYHPKYSDADIANPVNDTDWYDRITGTGQHTNISINGTEYTKTLISGNFF 447
QY	355 DQSGMAREPANPKRYSGRNFESRINELWLGANLSCAIANRRSADYFGKYVMGSGT--- 411
DB	448 NQGI VKN-NGMSRYTGRVNLQKLSYAKVGINLT---VSRNTLD---NVPLGAGONEY 500
QY	412 FGVLTPRYNPF----DVNGDLADVYMYGATRPSMTBEPYFAKMRPFSESHQAN--VN 465
DB	501 ASILVSAAQFSPLLSVKDENG-----YSLNQAAAYIPNPVSLLEISDQTTKERFLAT 553
QY	466 GFAQITPIKGLTLKAAGVDTITNRTSSKMPNPNPYDSTPLGERRE-RA-----YRDVSKS 520
DB	554 PFVEIKPINELTKASFGID-RNYQRREVYMPK-----TTLYGSKAGRADIGQYDRSDYL 608
QY	521 FTNTAEYKFSIDEKHDLTALMGHEYIEYEGDVI GASSKGPESDKMLLSOCK-----T 573
DB	609 LEITANYAKRLGD-HNLNALVGSYFRTFSKYLNAGNQGLTDAFLFNNL GACTYKPKW 667
QY	574 GNSLSLPEHRVAEYAYLSFFSRPNYGDKWYIDFSVRNDQSSRFGSGNNRSANFYSV--G 631
DB	668 GSSASKSE-----MASFFGRVNYTYKDRYLVTATLRADGASNFAKNNRWGFFPSVALG 720
QY	632 GMFDIYNKIQBSNW---LSDLRKMSYGTGNGSEICNYNHQALVTNNY---TEDAMGL 685
DB	721 WRFTEEN-FARSLNLDKVLNSGKRLSYGQTGNSNTGDKSVTYGYGYNKVFNGKEYTGV 779

ID	Q33PC3	PRELIMINARY;	PRT; 1089 AA.
AC	Q33PC3;		
DT	01-DEC-2001 (TrEMBLrel. 19, Created)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)		
DE	MS106, putative outer membrane protein.		
OS	Microscilla sp. PRE1.		
OG	Plasmid PSD15.		
OC	Bacteria; Bacteroidetes; Sphingobacteria; Sphingobacteriales;		
OC	Flexibacteraceae; Microscilla.		
OX	NCBI_TaxID=155537;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=PRE1;		
RA	Zhong Z., Toudarian A., Helinski D., Knauf V., Sykes S.,		
RA	Wilkinson J.E., O'Bryne C., Shea T., De Loughery C., Capi R.;		
RT	"Sequence Analysis of a 101 kb Plasmid Required for Agar-degradation		
RT	of a Microscilla strain.";		
RL	Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.		
CC	-1- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).		
DR	EMBL; AF339846; AAK6282.1; -.		
DR	InterPro; IPR00531; TonB_box.		
DR	Pfam; PF00593; TonB_dep_Rcc; 1.		
KW	Membrane; Outer membrane; Receptor; TonB box; Plasmid.		
SK	SEQUENCE 1089 AA; 120064 MW; 3EGD405FDG27DBE5 CRC64;		
Query Match 17.1%; Score 904; DB 2; Length 1089;			
Best Local Similarity 27.6%; Pred. No. 3.9e-39;			
Matches 306; Conservative 184; Mismatches 460; Indels 158; Gaps 42;			
Qy	2 KRMTLPFLC-----LLTSIGWMAQNRTVKTGTVISDNEPLIGANVVVGNNTTIGAATD 56		
Db	48 RRITLFPKCVALLLLPLANLSFAQVR-VTGTVTDDADDGSGLPGVSVVVG-TAQGTVTD 105		
Qy	57 LDGNFTLSVPANAKMLRVSVSGMTTKEVAL--ANVMKVLDPDSKVLQEWVVLGYGTGQK 114		
Db	106 IEGNTYTLVSQESDATALPFSFVGYLTQNVQLNGRSSVDLSKLDIQQLDEVVVVGYV-QK 164		
Qy	115 LSTVSGSVAKVSSKLAEKPVANTMDALQGVAGQVMTTSGDPTAVASVEIHGTGSLGA 174		
Db	165 KKEVTGAVGSIKSELKNATPDIGDAWGQIAGVNVQASGRPGAKSNIQIRIGISLSP 224		
Qy	175 SS-APLYIVDGMQTSLDVATMPNDPESVMSVLKDSATSITYGARAANGVVFIOTKKGKM 233		
Db	225 SGLGPLVVDGVPYQND--PNIAPEQVESIEVLKDGAAAIYGVRSNGVLIITTKGKA 282		
Qy	234 SERGRITFNASYGISQILNTKPLDNMTGDBELLFQVKAGFWGNQTVQVKQMDILAGAE 293		
Db	283 GEM-KVDSAYKGIQNTISGTPL--MNTQQMYYEEVKTEALGOEPQIFRETPDALDYNT 339		
Qy	294 DLVGNYSLXDYEKTLFPVDVFNHDAWLKALPKTAPTSGDISFSGSGSQTSYASIGY 353		
Db	340 DFVG-----DVQND-----NAAMQNTLVNMGKGBELTFSMNAY 374		
Qy	354 FDOEGMAREPANFKRYSGRLNFESRINEWLKVGANLSGAIANRRSADYFGKYMGSGTFG 413		
Db	375 FQDGVLIN-SGDFRFTRLSGEYKRNPF-KAFATVANTEENTQEWP-----A 421		
Qy	414 VLTWPRYNPDPVNGDLADVVMYGATRPSMTEPYPAKM-RPFSSSE-SHQANVNGFA--- 468		
Db	422 LYELAIYVQKEPWQRINDLESFGAGVLVPSDNPIDFGLSRQLNNEKDKRVKNNNNIALNL 481		
Qy	469 --QITPTIKGLTLKAAGVDITNTTSSKMPNNPYDS----TPLGERRERAYRVDKSFT 522		
Db	482 SYEFLP--GLSYQVNLGRNSWNYQRKFPPOYLAYSADGGIOPAGSNMD-ALLDESYYTF 538		
Qy	523 --NTAE--YKFSID-EKHDLTALMGHEIYIEGVGVIGASSKGFESDKMLLISQKTKGNSL 577		
Db	539 TKNTLENIVKFORNFGKHDLGATAVLSYENFNSKNVGVGVGLLSNDTDLVSAGTYGK- 597		
Qy	578 SLPEHRVAEYAYLSFFGRNFYGFQKMYVIDFSVRNDQSSRFSGNNRNSAWFYSGVGMFDIY 637		
Db	598 --PTGTSVQNIIGKMARVOYSDSRVLYFASIRYDGSVSPGKENVNYPFGIGISAGNLS 655		

Qy	638	NK-FIQE---SNWLSDLRKMSVGTGTGNEIGNHNQALVTVN---NYTDEAMGLSISTA	690			
Db	656	EEGFFONIGMLNFVSALKIRVSYAEALGNOSIGPY--QAVASIESGIDIPYPGEGABYLGV	713			
Qy	691	G-----NPDLSWKOSQNFGLAAGAFNNRLSAEVDVFVRTTNDMLIDVPVPYISGFF	743			
Db	714	GAIQRRPANPLIQWETTISRNIQIDWGMLEDRULUTADFYLNDDKMDMLAERLAPSGTW	773			
Qy	744	S-----QYQNVGSMKMTGVDSLKGTIYQNK-----DMNVYASANFNYNR-----QE	785			
Db	774	QTRAVGVYNNVRTINAGNMQNGVEVALG---YRDQTSFGLQWSVNGTFTTNVNRVTDLNG	830			
Qy	786	ITKLFLGINKYMLPNTG-----TIWEIGYP-NSFYMAEYAGIDK-----	823			
Db	831	IEGIAFGGGR-PITSSGERTDFTYLSQGYEGGAFFLEHIGVIKTEQLAAVREVEPSA	889			
Qy	824	KTGKQLWYYPGVQVDAQGNKVTTSQYSA--DLETRIDKSVTPPTITGGFSGLASWKGLSLDA	881			
Db	890	MMGDMMYRDLQTVDDTDGDGIPDAGDGLINDDDRVYAGSGQPEFEAGLNUNGA YKGFDLFI	949			
Qy	882	DFAYIVGKWMINNDRYFTENAGGLMQLNKDKMLLNAAWTEDNKETDVPKLGQSQPFD---	937			
Db	950	QMYYSHGAEIYNGAKLYAVGTG-----RHRDLYYMWSPQNPDSIPTARTSQHNNFRA	1003			
Qy	938	--THLLENASFLRLKNLKYTVLPNSLFAQQNVIGGARVYLMARNLLTVTKYIKGFPDEAG	995			
Db	1004	RSNYFLGDTGYLRIRITLGYALPSNLFGSK--VDNVRFPYITAAQNPFPTFTKYEYDPEIG	1061			
Qy	996	GN-----VGKNQYPSNKQYVAGIQLSF	1017			
Db	1062	GDGLFTRGVGVNGYPIYTRKFLGGIQVKF	1089			
RESULT 5						
Q93FV1	ID	Q93FV1	PRELIMINARY; PRT; 947 AA.			
AC	Q93FV1					
DT	01-DEC-2001	(TEMBLrel. 19, Created)				
DT	01-DEC-2001	(TEMBLrel. 19, Last sequence update)				
DT	01-MAR-2003	(TEMBLrel. 23, Last annotation update)				
DE	TonB-linked outer membrane protein.					
GN	OMP.					
OS	Bacteroides cacaee.					
OC	Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;					
OC	Bacteroidaceae; Bacteroides.					
OC	NCBI_TaxID=47678;					
RP	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=21437605; PubMed=11553542;					
RA	Wei B., Daliwadi H., Gordon L.K., Landers C., Bruckner D., Targan S.R.,					
RA	Braun J.,					
RT	Molecular Cloning of a Bacteroides cacaee TonB-Linked Outer Membrane					
RT	Protein Identified by an Inflammatory Bowel Disease Marker Antibody.",					
RL	Infect. Immun. 69:6044-6054(2001).					
CC	-1- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).					
DR	EMBL; AF305878; AAL09385.1; -					
DR	InterPro; IPR000591; TonB_boxC.					
DR	Pfam; PF00593; TonB_dep_Rec; 1.					
KW	Membrane; Outer membrane; Receptor; TonB box.					
SQ	SEQUENCE 947 AA; 105182 MW; F702662A38B4A2E1 CRC64;					
Query Match 15.9%; Score 843; DB 2; Length 947;						
Best Local Similarity 26.9%; Pred. No. 5.2e-36;						
Matches 279; Conservative 175; Mismatches 416; Indels 168; Gaps 40;						
Qy	57	LDGNFTLSVPANAKMLRVSYSGMTTKEVATA--NVMKIVLDPDSKVLQVQVVLGYGTGQK	114			
Db	1	MDKYSISVTPD--DVLAFSIVGMTSSVKVGAQTIVNVLKESQVLAETVVIYGSAKK	59			
Qy	115	LSTVSGSAVKVSEKLAKEKPVANIMDALQSQVAGMQVMTTSGDPTAVASVEIHGTGSLGA	174			
Db	60	-RLDTSGITNKGEEELANKPAMPLSLOQKAGVOIV-NSGRAGSDPEIRIRGTNSING	117			

QY 175 SSAPLYIVDMQTSLLVATMPNDPESMSVLKDSATSIIYGARAANGVVFOTKKGKMS 234
 DB 118 YK-PLYIVDGLFN--DNINFLNPEDIESMEILKDPSSLAIFGVRGANGVILITTKKKEG 174
 QY 235 ERGRITFNASYGISOILNTKPLDNMTGDELLDFQVKGFWGNQTVQVKDMILAGAE 294
 DB 175 QT-LVNINTSFGKVKVVKLVNGSQFRELSEQ-----LANQBD 214
 QY 295 LYXNYDSLKDEYKTLFPVDF--NHDADWLKALFKTAPTSQODISFSGSQTSYASI 351
 DB 215 A-----PDTFGWANTDQDEIFQTAFTNNISITGASPKHSFYLG 258
 QY 352 GTFDQEGMAREPANFRYSGRLNFBESRINEWLKVGANLSGA-----IANRSADYF 402
 DB 259 GSYEQGNI-EHEKFSKVITINASNDYKITDYLKVGFGQFNGARMLPADSKVLNLRATPI 317
 QY 403 GKVYMGSGTGGVLTMPRYNPDVNGDLADVYMYGATPSMTEPYFAKMRPFSESHQA 462
 DB 318 APVY--NDEYGLYSALPEFOKAQINPMVDVSLRANTTK-----AENYRA 360
 QY 463 NVNGFAQITPIKGLTLKAQAGVDITNTRTSSKRMNNPYDSTPLG-----507
 DB 361 SGNIVGEVDFLKHFNKAMFSDYASNNGRITQPIVKVVDPTVSGNIATLGTGKTEVSQF 420
 QY 508 -ERRERAYRDVSKSFNTAAYKFSIDEXHDLTALMGH-EYIEYEGDVIGASKGF-----560
 DB 421 KENETKVQSDYVLTNTNSPD-----NGENLTATAGFTTYNLSRLDGRKQGVGLVLP 475
 QY 561 -BSSDK-LMLLSQKGTGNSLSLPEHRAEYAYLSFSPRFNYGPDKMYIDFSVRNDOSRP 618
 DB 476 NPDKMFVSGDAATNGSTOWER-----STLSMLARVIYNYGKYLFNGSPFRDSSAF 531
 QY 619 GSNRSAM--FYSVSGMFPI-YNKFTQESNWLSDRLKMSYGTGTGNSEIGN-YNHQALVT 674
 DB 532 -SYTGNEMQNFPSLGGWLMSEEFWKDKLWMLKIKASYGTGLNQNLDKAYPAEPLLS 590
 QY 675 VNNYTEDAM-----GLSISTAGNPDLSEKQSFNFGLAAGAFNNRLSAEVDYVR 725
 DB 591 -NAYS--AVFGKPSIITPGVQLAYLPNPLRWEKVEAWGAPETNLLRNLHFEYGYWK 647
 QY 726 TTNDMLIDVMPYISGFFSQYQVNGSMKGTGVDLSLKGTYIQNKQWVYASANFNROE 785
 DB 648 NTKDLAEY--PGISITFGIGNLQIQNKGVEMAVTWR-DQIGENGYSVANLTTIKNE 704
 QY 786 ITKLF-FGLNKYMLPNTGTIWEIGYNSPYMA-----EVAGIDKTKGKOLWYV-PQ 835
 DB 705 VKSLVQEGYSIAGDKQSQYTWAGYIGFYGYKVEGVYQSQADIDASPKNTLATVTPGD 764
 QY 836 V---DADGNKVTTSQYSADLETRIDKS-----TPPITGGFSLGASWKLGLSLDADFAIYV 888
 DB 765 LKPADVDGNGEITPE-----DRTMIGNPTPKVYTGFSLGNYKNWSIGID---MMG 812
 QY 889 KWINNDRYFTENAGGLMQLNKKMLNNAWTEDEKTDVPKLGQSPQD-----THLENA 944
 DB 813 QG--GNKIERTDNYNFAQNFYLEQRLDRWHGEGTSNTQPLNLSKHSINLYNSDYIESG 870
 QY 945 SFRLKLNKLTVLPNLSFAGQVIGGARVYLMARNLLTVTKGPDPEAGG-----NVG 999
 DB 871 NFRIRINVLAVAFDKNLL-GKIRLQALKVYVNIQNLKTKWNTGYTPELGGTATAFGVD 929
 QY 1000 KNOYNSKOYVAGIQLSF 1017
 DB 930 NGSYPVPAVYTGFINLP 947

RESULT 6
 Q93TH9 PRELIMINARY; PRT; 1125 AA.
 AC Q93TH9
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)

DE Outer membrane protein Omp121.
 OS Bacteroides fragilis.
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Bacteroidaceae; Bacteroides.
 OX NCBI_TaxID=817;
 RN [1]
 SEQUENCE FROM N.A.
 RC STRAIN=ATCC25285;
 RA Wexler H.M., Read E.K., Tomzynski T.J.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
 DR EMBL; AF357210; AAK38604.1; -;
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_dep_Rec; 1.
 KW Membrane; Outer membrane; Receptor; TonB box.
 SQ SEQUENCE 1125 AA; 124650 MW; 05A9F2D5D67856C3 CRC64;

Query Match 13.9%; Score 734.5; DB 2; Length 1125;
 Best Local Similarity 23.6%; Pred. No. 3.4e-30;
 Matches 284; Conservative 218; Mismatches 428; Indels 271; Gaps 49;

QY 1 MKRMTLPFLCLLTISIGWMAQNRVTGTVISSEDENEPLIGANVVVVVNTTICAATDLQCN 60
 DB 12 MKKVLVLVLSLSVT--APAQNIIVKG-IVKDGCTGEPHIGSVLVKG-SSIGTVTDVGN 67
 QY 61 FTLS-VPANAKMLRVSYSGMTTKEVAIA--NVMKIIVLDPDSKVLBQVVVLGYGTQKLS 117
 DB 68 YTLNVPADG-VLEPSYIGMKKQDVKSGKTIVNVVLEQDTQILDEVVVVLTALGLKREOKA 126
 QY 118 VSGSVAKYSSEKLAEPVANIMDALQOVAGQVMTTSGDPTAVASVEIHGTGSLGASSA 177
 DB 127 LGYATEYKGGDGLKAAANTISFVAALQGVAGVEIRQSDGGMFGATKIQIRGASTLKGNNQ 186
 QY 178 PLYIVDGM-----QTSLD-----VATMNPNDPESMSVLKDSATSIVG 216
 DB 187 PLYIVDGVILNDSGTNTMDWAGNNANDYGNELKNLNPDDFTSVLVKGAATATLVG 246
 QY 217 ARAANGVYFIQTKGKMSERGRITFNASYGISQILNT-----KPLDNMT 261
 DB 247 SRLGANGAVIITKSGKFGFVSVQTFGIDHAYRTDPTQIEYGVGLMPGKWDN--N 304
 QY 262 GDELLDPQVAGFWGNQTVOKVKMDILAGAEADLYGNVDSLKDEYKTLFPVDFNHDADW 321
 DB 305 GSVMPDFQKLDKGD-----RTLIGAGSYGMPKYDQG-----PIR-NVDGTW 347
 QY 322 -----LKALFKTAPTSQDISFSGSQTSYASIGYFDQSGMAREPANFRYSGR 372
 DB 348 TNYSPHKNNMLDLYQLGLNSNTNVAIRGGNDKTSYITSLSY-KKARSTSEKNTFFERYSPL 406
 QY 373 LNFESRINEWLKVGANLSGAIANRRA-----DYFGKTYMGSG 410
 DB 407 LKSGHKISDRVEVSAAMSPFTNSPKNSPRTVGERFVNPNGTITPMLDVNYPDKYLGE- 465
 QY 411 TFGVLTPRYNPP-DVNGDLADVYMYGATRPSTPEPYFAKMRPFSESHQANVNGFAQ 469
 DB 466 -HGGLASTSYGDKYGSVPG--RDLFFMDIKDYDSKE--TVVRP-----QMEVN-----509
 QY 470 ITPKGLTLKAQAGVDITNTRTSSKRMEN--NPVDSSTPLGE-RREAYRDVSKSFTNTA 525
 DB 510 VOILDWLRFKADANNYYTTFPEEKQLGSGYANEGKGTMTGQTTKEQA-----TFGGT- 562
 QY 526 EYKFSIDBKHDILTALMGHEIY-----EGDIVGASSKGFSDKMLLSQKGT 573
 DB 563 ---FTVNKOIQDFSVGFGFARVEYITSRSEAYKYVTDGGVVGQVFWFVNSKNPKSEASI 619
 QY 574 GNSLSLPEHRAEYAYLSFFSRFNYGFDKMYIDFSVRNDOSRRP-----GNNRSAMFYS 629
 DB 620 SNTKEM-----MSAVPALNLGKVNQVLDVTGRNDWSSSLVYQNGMGTYSYFYPS 669
 QY 630 VCGMFDIYKFTQESNWLSDRLKMSYGTGTGN-----SEIGNYHQAALVTNNY--TE 680
 DB 670 VSGSW-LLNETFDLPHWITPAKVRGSAQVGNQDTPDYVYNSVYGFETKEMDGNIVNTL 728

```

QY 681 DAMGLSISAGNPDLSWEKOSQFNGLAAGAFNNRLSAEVDYFVVRTNMDLIDVPMPEYIS 740
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
729 D-----KTKSLKLPKRNKAWVEGLDLRLFDLSRLNFDFTYKENTRDQIMSLVPAIS 782
QY 741 GFPSQYQNVGSMKNTGVDLSLKGTYIQNKQNVNVSANFNYNROBITKLPFGLNKMYL-- 798
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
783 GVNQLINAGNIQNGIEIAVNATPYKNQKQWQVAMTYTKNKNITISLHENVADYIALS 842
QY 799 -----PNTGTIWEIGYPNSFYMAEYAGIDKKTGKOL-----WYVPGQVDADGNKVT 845
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
843 GYANDYDHIGSVAGQDGYGLLMSDILPAKNEKGETLLEWDSW--RGAYEARSQGV-- 898
QY 846 SQVSADLETRIDKSVTPITGFGSLGASWKGSL-----DADFAYIVGK-----WM 891
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
899 -----QEVGK-WTDFLGLSLTTLNKNLSLHIATDMRFGGLVASYSNLYCTQAGWI 949
QY 892 INNDRYTENAGGLMQLNK-----DKMLLN---AWTDNKETDVPKL--- 930
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
950 KSSLKWRDPBHGGLSWTSQYGDGKISYGVGCVIPDGVPKNGTATLVDTGKMDVSGMSYK 1009
QY 931 -----GOSQPDTHLLENASFLRLKNLKLTYVLNPSL-----FA 964
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1010 QLVAGKLEPHTAGTYHVNRAAWGNTFTDFTWHE--LNYIALREITISYRPPKSVASKFG 1068
QY 965 GQNVIGGARVYLMARNL--LTVTYKGFDPPEA--GQNVG-----KNQYPSNKQYVAGIQLS 1016
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1069 AQ-----GLGLSFSARNLGYLNSLPNHLNPSVGRGNTASEFRIRGYEPTYANYMTINVD 1124
QY 1017 P 1017
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1125 F 1125

RESULT 7
Q9AF86
ID Q9AF86 PRELIMINARY; PRT; 1089 AA.
AC Q9AF86;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Outer membrane protein Omp17.
GN Omp17.
OS Bacteroides fragilis.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=817;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC25285;
RA Read E.K., Tomzyński T.J., Wexler H.M.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR EMBL; AF364409; AAK35222.1; -
DR InterPro; IPR00531; TonB dep Rec.
DR Pfam; PF00593; TonB dep Rec; 1.
KW Membrane; Outer membrane; Receptor; TonB box.
SQ SEQUENCE 1089 AA; 120709 MW; 9C937302774101F0 CRC64;

Query Match 12.2%; Score 645.5; DB 2; Length 1089;
Best Local Similarity 23.4%; Pred. No. 1.6e-25;
Matches 248; Conservative 185; Mismatches 440; Indels 189; Gaps 37;

QY 21 QNRTVKGTIVSSEDNEPLIGNVVVVGNVTITIGATDLGDNFTLSVPAKMLRVSYSGMT 80
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
110 ESKLKG--IVKDDKGDPLIGNVVSGKSGPT--GTVTGLDGRPSI--LAAKGNIIPEFVGYT 166
QY 81 TKEVAI--ANVMKIVLDPDPSKVLQVVLVYGTGQKLSVTSGSVAKVSSSEKLAEPVANI 138
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
167 TQYITVDASSLTVLEEDAKALDEVVVVYALGKRAEKALSYSVQVYKSDAINDVKDANF 226
QY 139 MDALQGVAGQVQVMTTSGDPTAVASVEIHGTGSLGASSAPLYIVDGMQTSI----- 189
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
227 VNLGTGKAGVGINRSGSGIGGATRVVMRGAKSIVGNVNVLYVVDGMPIGNPSKGEINND 286

```

```

QY 190 -----DVVATMNPNDPESMSVLKDSATSIYGARAANGVWFIOTKKGMSRGRITFN 242
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
287 YSTPGGEGISDPNPEDIESILITGPAAALYGSSAANGVILINTKGG--QEGKLKIS 343
QY 243 ASYGISQIILNTKPLDMMTGDLELDFQVKGAGFWGNNQTVQVKMDMILAGAEDLYGNYDSL 302
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
344 ISN-----NTE-----FMTPYVMPEFQNR--YGNAK-----GSYKSW 373
QY 303 KD--EYGKTLFPVDFNHDADWLKALFKTAPTSGDIDISFSGSGSQTSYYASIGYFDQEGMA 360
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
374 GEMLOQPSFTRPKDF-----FKTGANIMNAANFSGVGNKNQNTFFSVATTNSTGII 423
QY 361 REPAN--FKYSGR--NFESRINEWLKVGANLSGAIANRRSADYFGKYWGSGTGFVLTMP 418
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
424 --PNNYYRYNFTLRNTASWLNKHLN-----LGASYVLQGDQWLSAG 466
QY 419 RYNNP-----FDVNGDLADVYVYGATRPMSRTPPYFAKMRP--FS 456
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
467 RYFNPLVPLVLPFGEDFEAVKYERYDTNRKFPICQESYDQGLNLENPYMIWVNMRFV 526
QY 457 SESHQANVNGFAQITPIKGLTLKAQACVDITNTRTSK-----RMPNPNYDSTPL 506
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
527 SKKKRYMFYANVKYDILSWLNTRIAGRIRVDNTNTTSERKLHASTIKLHAOSDKGAYNRSM- 585
QY 507 GERRERAYRD---VSKSPTNTAEYKFSIDKHDLTALMHEYIEYEGDVGIGASSKGFES 562
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
586 -EYQOQYADIMLVNKNFCN-----FNLNAGSFSYEDHLTGMGIGGKLPV 633
QY 563 DKMLLL-----SQKGTGNSLSLPEHRVAYAYLSFFSRFNYGDKWYIDFSV 610
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
634 PNLFSAYNFDPASGPGSQSHTHTRNN-----SVFVSTELGYKSLMLYLTGTG 679
QY 611 RNDQSRFGSNRNSANFYSVGGMFDIYNKFIQESNWLSDLRLKMSYGTTCNSIGNYNHQ 670
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
680 RQEWASQLVNSDQPTYFYPFSGVSGVISEMISLPKFSIFWKMRAEPAEVG---GPINYT 735
QY 671 ALV--TVNNYTEDAMGLSISTAGNPDLSWEKOSQFNGLAAGAFNNRLSAEVDYFVVRTN 728
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
736 GLTPGVTPDMKGVNINPISVYPPFNPKAEQTKSYELGTNLRFLFNKNINIDATVYLTDTY 795
QY 729 DMLIDVPMPIYSGFSSQYQNVGSMKNTGVDLSLKGTYIQNKQNVNVSANFNFY--NRQEI 786
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
796 NQTFLLSSMSPASGYSGYVQAGVKRNKGIELSLG---YNDRFKGVGYATNLTYTANRNI 852
QY 787 TKLFFGLNKYMLPNTGTIWEIGYPNSFYMAEYAGIDKKTGKQLM--YVPGQV--DADGNK 842
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
853 MKM---VHDYKNPSDGLFSI---TELTLDQKGGVYLREKDAIGDYYVVGILARGKDGKL 906
QY 843 V-TTSOYSADLETRID--KSVTPPTITGGPSLGSAGWKLSDADPFIYVGMKMINNDRYFTE 900
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
907 IEIEGGYKVDRSQRIKIGSVNPDPSIGWRHNVVNNITDILLFNGRFGGIVTSSTQAFLD 966
QY 901 NAGGLMLKNDKMLLNW---TEDNKETDVPKLQSQPQFDTHLLENASFLRLKNLKLTYV 957
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
967 DYGVSKDTVKARQGVVWNGIQYDAEKYVTTIG--GEQLMAYVAYKATNIRLQEASLSYT 1025
QY 958 LPNSLFAGQNVIGGARVYLMARNLLTVTKYKGFDPPEAGNVG 999
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1026 LPQKWFQ--NVINRLTVSAIGRNLMLWMIYNKAPDPPEMTSTG 1065

RESULT 8
Q9XAU9
ID Q9XAU9 PRELIMINARY; PRT; 59 AA.
AC Q9XAU9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Raga protein (Fragment).
GN RAGA.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;

```

OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_taxid=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=W50;
 RX MEDLINE=20316027; PubMed=10858216;
 RA Bonas W.A., Marsh P.D., Percival R.S., Aduse-Opoku J., Hanley S.A.,
 RA Devine D.A., Curtis M.A.;
 RT "Identification of ragAB as a temperature-regulated operon of
 RT Porphyromonas gingivalis W50 using differential display of randomly
 RT primed RNA.";
 RL Infect. Immun. 68:4012-4017(2000).
 DR EMBL; AJ242673; CAB46019.1; -;
 FT NON_TER 1 1
 FT 59 59
 SQ SEQUENCE 59 AA; 6486 MW; D3527513DABR333AC CRC64;
 Query Match 5.7%; Score 302; DB 2; Length 59;
 Best Local Similarity 98.3%; Pred. No. 2.3e-09;
 Matches 58; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 429 DLADVYMYGATRPSTPEYFAKMRPFSSSHOANVNGFAQITPIKGLTLKQAQGVDT 487
 Db 1 DLADVYMYGATRPSTPEYFAKMRPFSSSHOANVNGFAQITPIKGLTLKQAQGVDT 59
 RESULT 9
 Q8KB41 PRELIMINARY; PRT; 844 AA.
 AC Q8KB41
 DT 01-OCT-2002 (TremBLrel. 22, Created)
 DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)
 DE 01-WAR-2003 (TremBLrel. 23, Last annotation update)
 DE Ferric siderophore receptor, putative, TonB receptor family.
 GN CT1953.
 OS Chlorobium tepidum.
 OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
 OC Chlorobium.
 OX NCBI_TaxID=1097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TLS / ATCC 49652 / DSM 12025;
 RX MEDLINE=22103685; PubMed=12093901;
 RA Eisen J.A., Nelson K.B., Paulsen I.T., Heidelberg J.F., Wu M.,
 RA Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,
 RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
 RA Holt I., Unayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
 RA Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,
 RA Vamathevan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,
 RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
 RT "The complete genome sequence of Chlorobium tepidum TLS, a
 RT photosynthetic, anaerobic, green-sulfur bacterium.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
 DR EMBL; AS012945; AA073172.1; -;
 DR TIGR; CT1953; -;
 KW Receptor; Complete proteome.
 SQ SEQUENCE 844 AA; 93821 MW; E2C0ED014B0E7B8A CRC64;
 Query Match 5.1%; Score 270; DB 16; Length 844;
 Best Local Similarity 21.2%; Pred. No. 6e-06;
 Matches 224; Conservative 139; Mismatches 368; Indels 328; Gaps 54;
 QY 1 MKRMTLFFLCILTSIG----WMAQNRTVKGTVISSEDNNEPLIGANVVVGNVTTGAATD 56
 Db 11 MKRLASATFLATLAPVTW--GADGVIRGRVTDKADGEGVGVGAIVSAG--TNATATVD 67
 QY 57 LDGNFTL-SVPANAK----MLRVSYSGMTTKEVAIAN----VMKIVLDPDSKVLQVVVL 107
 Db 68 INGNFVLRNPASKQKVIVTSIGYA-PTQVINLGDGQTATINIALGQTTIMASEVVVG 126
 QY 108 GYGTGQKLVSGSVAKVSEKLAEPVANIIMDALQGQVAGVQVMTSDGPTAVASVEIH 167
 Db 127 AALYKQDRLDVFPVTANIVTKEQIREEPNPTLDEVQ--DVFGVVVSRAAG--TSSSNLQIR 183
 QY 168 GT-----GSLGASSAPLYIVDMQ--TSLD-----VVATMPNDFESMSVLKDSATSIYGA 217
 Db 184 GSNYNGGGIGTRVNAFY--DGFPIPSDGSBIVQSVNMAADKVELKGAAT--LYGS 240
 QY 218 RAANGVFIQYKKGMSERGITFNASVIGISQILNTKPLDNMTGDELLDFOVKAGFWGN 277
 Db 241 GAMGGVNNI-----TGHLDPKFEVFKAG----- 262
 QY 278 NQTVQKVMILAGAEDLYGNYD-----SLKDBYEGKTLFPVDFNHDADWLKALFKTAPTQS 333
 Db 263 -----SG-----IGFYDKTPSSDESEYRGTPTVFWNTYA----- 292
 QY 334 GDIFSGSGQSTSYVASIGFYDQEGMAREPANFK---RYSGRINFESRINELWKYGANLS 390
 Db 293 ---GFGNKSQKWTYDLYSHSDDDGYRQVWYNNYMDVKFARYDIDSR----- 337
 QY 391 GAIANRRSADYFGKYMGSGTFCVLTMPRYNPPVNGDLADVYMYGATRPSTPEYFA 450
 Db 338 -----QY-----LQLTSFYN--STVGYA--YOWPYNATISTST----- 367
 QY 451 KMRPFSSSHOANVNGFAQITPIKGLTLKQAQGVDTITWRTSSKRMNPNPYDSTP-LGR 509
 Db 368 -FTPILDQSYDVFN--ARLFP-----THTAAPSIAQTMYPSTPFLMDI 408
 QY 510 RERAYRDVSK-----SFTNTAEYKFSIDK-----H 535
 Db 409 LANAWSTYTKVDVYTDLISRKNALVGINYVNLSSDKLSLDRLYYYTNASRIEYNRTDA 468
 QY 536 DLTALMGH-----EYIEYEGDVIGASSK--GFESDKMLLSQGTGNSLSLPEHRVA--BY 587
 Db 469 DQYATGIRITIGEFNETDSDRYGAGIKLDWRASNHRLLF--GVGDNIVDTRTTQVAVEY 527
 QY 588 AYLSPFSRF---NYGF---DKWYID-----PSVENDOSSRRSGSNRSANFYSVGMFPI 636
 Db 528 PVKNEFNIOEKNFAVFLQDEWKITDKLTSLSLRYDWS---GVNKEVEITPGWPIPI 583
 QY 637 YNKFIQ-----ESNWLSDRLKMSYGTTCGNSEIGNYNHQAALVTVANNYTBDMGLSTS 688
 Db 584 NKKSVDALSPRVALNYRATDDMALRASNGR-----SPRAPSLYERFVHDAGFLTV- 633
 QY 689 TAGNPDLSEKQSQNFGLAAGAFNNRLSAEVDVFTVTTNDMLIDVMPYISGPFSSQVON 748
 Db 634 -VPNPDLDKETMTAWEAGIFK-QFSDKVSLD-----IAGFINNYDN 672
 QY 749 VGSMTKTCVDLS-LKGTIYQKNQVYASANFNYNRQETIKLFFGLNKYMLPNTGTIWEI 807
 Db 673 LIESPTAAPLYMYGNITKARINGI--ETNLNVR-----PNTDNLVS 714
 QY 808 GYPSNFYMAEYAGIDKKTG-----KQLWYVPGQVDADGNKVTTTQVSADLETRID 857
 Db 715 GY--TYMAKNRSFDASTATATELNNPDPEWLPYRPEHT-----ASASTWKAT 761
 QY 858 KSVTPPIITGSPSLGASWGLSDADFAIVYGVKWMINNDRYFTFENAGGLMQLKMKLNA 917
 Db 762 KKLTLNVNGRYV--GKYKAVTL---YTNPDGKWYPCD--FVVFNAGLKYQFNKNVTATLA 814
 QY 918 WTEDNKETDVPKLGSPQDFTHLENASFLRLKNLKIT 956
 Db 815 CNNINNT-----QYSEAEWFRAPNRSF--IAGIDLTY 844
 RESULT 10
 Q9KIBO PRELIMINARY; PRT; 757 AA.
 ID Q9KIBO
 AC Q9KIBO
 DT 01-OCT-2000 (TremBLrel. 15, Created)
 DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
 DT 01-WAR-2003 (TremBLrel. 23, Last annotation update)
 DE Hypothetical tonB-linked outer membrane receptor Pgl3.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.

OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B., Barr I., Patterson M., Agius C., Rothe L., Margetts M.,
RA Hocking D., Webb E.;
RT "P. gingivalis" polypeptides and nucleic acids";
RL Submitted (PEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B.C.;
RT "Genomic analysis of Porphyromonas gingivalis for vaccine discovery";
RL Submitted (PEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR EMBL; AF237558; AAF81416.1; -;
DR InterPro; IPR00531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rec; 1.
KW Membrane; Outer membrane; Receptor; TonB box.
SQ SEQUENCE 757 AA; 84517 MW; 224B6D65264F9D62 CRC64;

Query Match 4.4%; Score 232.5; DB 2; Length 757;
Best Local Similarity 20.2%; Pred. No. 0.00048;
Matches 171; Conservative 124; Mismatches 279; Indels 271; Gaps 41;

QY 1 MKRMTLF-----FLCLITSGWMAQNRV-KGTVISSEDENEPLIGANVVVGNNTTIGAA 54
DB 1 MKRTIIFPAISIFALLUSS--LSAQKAVLTGSSVDAETGEPLAGAR-IEVXHNIVAG 57

QY 55 TDLGDNFTL-SVPANAKMLRVSYSGMTTKEVAIA-----NVKMLVLDPPDSKVLQVQVVL 107
DB 58 ADAGGHEFEIKNLPAQOHTIICSLGGYQKEEVAIEAGQTKTISFALRLTNLEEUVVT 117

QY 108 GYGTGQKLTSSGVAKVSEKLAEPKVPANIMDALQGVAGQVMTTSGD--PTAVAS-V 164
DB 118 GTGTRYRLVD-----APVATEVLTKADIASF--SAPTSEALLOGLSPSPDFGNLMGSEFM 170

QY 165 EIHGTGSLGASSAPLYIVDGMOTSLDV-----VATMNPDPFESMVLKDSATSIYGARA 219
DB 171 QLNLG-----SKYLIILIDKRVYGVGGQADLSRISPDQIERIELVKGAS--SLYGSDA 225

QY 220 ANGCVFIQTGKGMSEGRITFNASYGISQILNTKPLDNNMTGDELLDFQVKGAGFMGNQ 279
DB 226 IAGVINVTIK-----NTNRLSAYTSHRISK-----YNDRQ 256

QY 280 TVQVKDMILAGADLYGNVDSLKDEYKTLFPVDFNHDADWLKALPKTAPTSGDISFS 339
DB 257 T-----FYHTDQWNSPPEIKKKG--SGEPVLEETVK-----NTSLDINIG 266

QY 340 GSGQTSYASIGYFDOEGMAREPANPKRYSGRLNPFESRINEMLVKANLSGAIANRRA 399
DB 267 KPSNTNIF-----FYHTDQWNSPPEIKKKG--SGEPVLEETVK----- 305

QY 400 DYFGKYMGSGTGFGLTMPRIY--NPPDVANGDLADVYMYCATRPSMTPEYFAMRPFSS 457
DB 306 ----KTFRAGBNQVQSLSYATNNLSFSG--NVQY-----NKQIQTFTFESEKAYDM 354

QY 458 ESHQANVN-GPAITPTKGL-TLKAQAGVD-----ITNRTSSKMPNPNPYDSTPLGER 509
DB 355 DYRALTASLTGNVLFPP--NGLHTLSFDVYDFRFRGYLYHDKDSSESILNNQ-----GQT 407

QY 510 RERAY-----RDVSKSFNTAEYK--PSIDEKHDLTALMGHEVIEYEGDVIGASSKGFES 562
DB 408 EQPTFFPQQLRNNDQIRYTAERAGVFTLPYAKLGT--GLEYFREE----- 452

QY 563 DKMLLSQKGTGNSLSLPEHRVAYAYLSFPFRNFGYDFKWMYIDFSVRNDQSSRFGSNN 622
DB 453 -----LISPNYLIITDKADASTLSA--YVQDEWKPLD----- 481

QY 623 RSWAFYVGGMFDLYNKFIQESNWLSDLRKMSYGTGTSNSEIGNYNHQA----- 671
DB 482 ---WFNNTAGFRLVHHQ-----BFGTRMTPKVSIKAYGFLPNFRATYANGYKTPTL 529

QY 672 --LVTVNTYTEDAMGLSISTAGNPDLSWEKQSQFNFL-----AAGAFNNRLSAEV 720
DB 530 KELFARNELT--TMGSHNLYLGNADLKPMQSDYALGLEYNQGPISFSAIVYDNLRLNI 587

QY 721 DFVVRTNMLIDVMPMP-----YISGF--FSQYQNVGSMKNTGVDLSUKGTIYQKNQNVYA 775
DB 588 SF-----MDIPTSPHEAQQIKTKQYANIGKARSRLDVLCDASI-----GWIKL 634

QY 776 SANFN 780
DB 635 GAGYS 639

RESULT 11
QY1473 PRELIMINARY; PRT; 616 AA.
AC QY1473;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Probable tonB-dependent receptor.
GN PA1271.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber L.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.I., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen";
RL Nature 406.959-964(2000).
DR EMBL; AE004557; AAG04660.1; -;
DR InterPro; IPR00531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rec; 1.
KW Receptor; Complete proteome.
SQ SEQUENCE 616 AA; 67568 MW; F0C58A57622209A1 CRC64;

Query Match 4.3%; Score 226.5; DB 16; Length 616;
Best Local Similarity 18.8%; Pred. No. 0.00072;
Matches 158; Conservative 104; Mismatches 265; Indels 313; Gaps 32;

QY 79 MTTKEVAIANVMKIVLDPDSKVLQVVLGYGTGOKLSTVSGSVAKVSSEKLAEPVANI 138
DB 6 LTPAAVALCGASSLSLAEPVSLADQVVT-ATTAQTASQSLAAVSVIDREDI-ERSQARS 63

QY 139 MDALQGVAGMQVMTTSGDPTAVASVEIHGTGSLGASSAPLYIVDGMQ-----TSLDVVA 193
DB 64 VPELLRQVPGVS-LANNGGFGKNTTLFLRGT-----ESDHVLVLIDGIVKGSASAGLTAFO 118

QY 194 TNNPDPFESMVLKDSATSIYGARAANGVVFIOYTKGMSERGRITFNASVIGISQILNT 253
DB 119 DLPVELIERIEVVR-GPRSSLYGSEAIQVIFTRRGD-GQAKPFPFAGYGTGTHQ----- 172

QY 254 KPLDNNMTGDELLDFQVKGAGFMGNQTVQVKDMILAGADLYGNVDSLKDEYKTLFPV 313
DB 173 ----- 172

QY 314 DFNHDADWLKALPKTAPTSGDISFSGSQGTSYVYASIGYFDOEGMAREBPA----- 364
DB 173 -----TLEGSAGVSGAGNMGWYSLGVSSFDTAGINTKRACTAGVPEPR 215

QY 365 -NPKRYSGRLNPFESRINEMLVKANLSGAIANRSDYFGKYMGSGTGFGLTMPRIYNP 423
DB 216 DGYRNLNGLNGLGGRYFDN-----GLELDGTLRAKSHNDYDQVFGNSG----- 258

QY 424 FDVNGDLADVYVWYGATRSMTPEYFAKMRPFSSSHQANVNGFAQITPIKGLTLKAQAG 483
 Db 259 FANAD-----GEQNLVGRARFPFPDPLVTLQAG 289
 QY 484 VDIWTRTSSKMPNPNYDSTPLGERBRAYRDVSKFTNTAEYKFSIDKEDLTALMGH 543
 Db 290-----RSEDKADAYD-GRFYRFDTRDLSLQNDLTLAEGH 326
 QY 544 EY-----IEYEGDVIGASSKGFESDKMLLSQGTGNSLSLPEHRVAEYAYLSFSSRF--NY 598
 Db 327 VLTGLGYDWQDEI-SSSEAFSVDS--RLNKG-----WFAQVLCGY 363
 QY 599 GFDKMYIDFSVRNDSSPFG--SNRSASFVSVMGMDIYNKFTQESNWLSD-LRLKMS 655
 Db 364 GRQDW--QLSLRDNNOQFVHDGTGSAANGYA-----LSDALRFTVS 403
 QY 656 YGTGTGNSSEIGNYHQALVTNNYTEDAMGLSISTAGNPDLSWEKQSQFNFL----- 707
 Db 404 YGTAFKAPFN-----ELYPDYGNPDLDATERSLEVGLSGTHGWGH 446
 QY 708 -AAGAFNNRSLAEVDYVTRTNDMLDVPMPYISGFSQYQNVGSMKNTGVDLSLKGTTIY 766
 Db 447 WAVNAFRTNV-----DDLIGNDPRPAPGRPWGPNNDIARIRGVELVL-GSQW 494
 QY 767 QKQNNVYASAFNVNQ-----EITKLPFGNL-----KYM 797
 Db 495 LGWMDNANATFLDPQNRSGGVNDGNELPARRRMFNLELDRRPERLSLGSVHABGRYD 554
 QY 798 LPTGTIWEIGYFNSFYMAEYAGIDKTKGQL---WYVPGQVDADGNKVTTSQYADLET 854
 Db 555 DP-----ANKVRLGGYATLDRSEYRLNDEWRLQGR-----ANLFGADHET 596

RESULT 12

QKIB1
 ID Q9KIB1 PRELIMINARY; PRT; 867 AA.
 AC Q9KIB1;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical tonB-linked outer membrane receptor PG47.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=W50;
 RA Ross B., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
 RA Hocking D., Webb E.;
 RT "P. gingivalis polypeptides and nucleic acids";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=W50;
 RA Ross B.C.;
 RT "Genomic analysis of Porphyromonas gingivalis for vaccine discovery";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
 DR EMBL; AF237557; AAF81415.1; -;
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_dep_Rec; 1.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Membrane; Outer membrane; Receptor; TonB box.
 SQ SEQUENCE 867 AA; 95231 MW; F8B4842E821C458E CRC64;

Query Match 4.3%; Score 225.5; DB 2; Length 867;
 Best Local Similarity 19.5%; Pred. No. 0.0014;
 Matches 207; Conservative 151; Mismatches 381; Indels 325; Gaps 55;
 QY 4 MTLFFCLLTSIGWMAQNRT-VKGTVISSENEPLIGANVVV-VGNTTIGATLDGNF 61
 Db 6 LQLFLTALLGSLATAQTAVTGTGKVIDSETSEPLIGSVSTGGQASLRGVTDTMDGGF 65

RESULT 13
 Q8PI21
 ID Q8PI21 PRELIMINARY; PRT; 1022 AA.

QY 62 TILSVPANAKMLRVSYSGMTTKEVALANVMK-----IVLDPDSKVLEQVVVLGY 109
 Db 66 RFEVPAK-----SVTFPCVGYATVTRSGRGSQEDLGTILLDPQAIGLDEIQVI-- 115
 QY 110 GTGQKLSVSGSAKVSKSEKLAEBKAVANIMDALQOQVAGMQVMTTSGDPTAVASVEIHGT 169
 Db 116-----ASYVPKDRMTPVPSNI-----RVADIQA----- 139
 QY 170 GSLGASSAPLYVDGMQTSLDVATMNPNDPESMSVLKADASATSIYGARAANGVVFIQTK 229
 Db 140-----ASLNV-----EPPELVK--STPSYITTKSGGPF-----G 166
 QY 230 KGMSEGRITFNASYGISQILNTPKPLDNMTG-----DELLDFQVKAGF----- 274
 Db 167 DORTNVRGFDTVN--FGV--LINGVPVNGMEDGKVYWSNWSGLMNAQASTIQIQRGLGASK 222
 QY 275-----WGNQTVQVKDMILAGAEGL-YGNVDSLKDBYKTKLFPVDFNHDADWLKALFK 327
 Db 223 LGISVVGTMNIIKTNTDANTGGSAYVMGN-----DGLHKSFSISTGMNDGWA----- 272
 QY 328 TAPTSQGDISFSGSGQGTSYVASICYFQDEGMAREPANFKRYSGRLNFESRINEMLKVGA 387
 Db 273-----ITTAG-----SHMTGLGYV--KGL-----KGRAPSYFFNVSKKFE--RHTL 310
 QY 388 NLSGAIA-----NRRSADY-----FGKY-YMGSCTGVLTPMPRYNPFVNGDLADVYMYG 438
 Db 311 SLTGFGAPQWENQRSSKYSVADYDKYGRHNOSFGYL-----RGELTPAYAYN 359
 QY 439 A-TRPSMTEPFAM-----RPFSSSHQANVNGFAQITPIKGLT 477
 Db 360 TYHKPQFSLNHFWMKDENTSLYAXYASLATGCGRRAYCKNSKWVLIN-YNTGQPYEQTK 418
 QY 478 LKAQAGVDITNTRTSSKMPNPNYDSTPLGERBRAY-----RDVSKSFNTAEYK-- 528
 Db 419 VTPDGLIDYDAVLAANAASNGSEAFALGNSNSHKWFGLLSFKKLLNSLTLTAGYDGR 478
 QY 529 -FSIDEXHDLTALMGHE-YIEYEGDVI GASSKGFESDKMLLSQGTGNSLSLPEHRVAE 586
 Db 479 YVRGHDYKIDTLGGSYIEDPKTKLAYHAEG-----QQLKVGDIYN--RDYTCGE 527
 QY 587 YAYLSFFSRFNYGDPKWM--YIDFSVRND--QSSRFGSNRNSAMFYSGVGMFDIYNKFIQ 642
 Db 528 IMWHGLFAQMEHS-SEWIDAFVSGSINVELYRNHNYGSKSGTGYLPV----- 574
 QY 643 ESNMLSDLRLKMSYGTTCNSEIGNYHQALVTNNYTEDAMGLSISTAG-----NPDLSWE 698
 Db 575 -SPWKSFLPWS--GKAGLSYKFAQGHNVFANGGFFTRAPLFGNIYAAGAILPNKANME 630
 QY 699 K--QSQNFGLAAGAFNNRLSAEVDYVTRTNDMLDVPMPYISGFSQYQNVGSMKNTG 756
 Db 631 KVLTEVCYG-----FTNHKPFEN-----INGYTKMMDRVTSKRIG 668
 QY 757 VDLSLKGTIYON-----KDMNVYASANFNYNQEIYKLPFGLNKYMLPNTGTIWEIGYNS 812
 Db 669 NEY-----VYLVGDAVHCGVEAEVSRPIRQIDLRGMFSLGDM-----TWQ-----NN 712
 QY 813 FYMAEYAGIDKTKGQLWYVPGQVDADGNKVTTSQYADLETIDKSVTPPTTGGFSLGA 872
 Db 713 VSYTSDYDEAGNETGQDITYIKGLHVGDAQAQMTAA-VSADIEL----- 753
 QY 873 SWKGLSLDAPYIVGKMMINNDRY--FTENAGGLMQLNKO-KMLNNAWTEDNKSTDPV- 928
 Db 754 -FKGF-----HVIGKYNFLGKNVAGPNPATRNAQQOYEADGKEIVESW-----KLDPVGL 801
 QY 929 -KLQSQPQDTHLLENASFLRLKNL-KLYTVLPNSLFPAGQNVIG 970
 Db 802 FDLASYNFKLSLSTTFYFNMDNVADKRYVSD-----ADDNIIG 841

Query Match	4.0%	Score 210	DB 16	Length 966
Best Local Similarity	20.6%	Pred. No. 0.01		
Matches	206	Conservative 139	Mismatches 345	Indels 312
Gaps	54			
QY	98	SKVLEQVWVLYGTCQKLSYV-----SSGVAKVSSEKLAEPVANIMDALQ--GOVAG--M 149		
DB	52	TQTLDYSVV-----TGSRLRRVDTETANPVVTVSQEQIAATGKATVGLQELPSIAGNAT 107		
QY	150	QVMTTSGDPTAVASVEITHGTGSLGASSAPLIVDGMOTSLDVTWATPNPDPEMSVLKDA-209		
DB	108	NPYTNNGGTCGASAVSRGLG-----DKRTLVLVNGIRLAYNDVNVAIPASMIERIEVLSD- 162		
QY	210	SATSIYGAARAANGVY--FIOTKKMKMSERGITFNASYGISQILNTKPLDNMMTGDLELD 267		
DB	163	GASAVYGSDAIGGVVNFILTR-----FDGVQFSSDFGTS-----KERDG-----NERN 206		
QY	268	PQVKAFGWGNQTVQVKVMDMLAGABDLYGNYDSLKDEYKTLPPV-----DFNHDA--- 319		
DB	207	FALTAG-----KTPE--RGSLLIAGLS--YHNID-----PVSAAARDYSKDALAL 246		
QY	320	DMLKALFKTAPTSQGDIFSFGSQ-----GTSYYVASIGYDQSGMA 360		
DB	247	TGDPVVKQGSATPTGTGVNFNDGSDASSQLRQANGCTRVTLNGVSGRAGPGDFHCYDAA 306		
QY	361	REPANFKRYS-----GRLNF-----ESRINELKVGANLSGAIANRRRSADYF 402		
DB	307	ADSNYQPENLQTPQKRTNAFLVLTGYRFSDSVSGYNTW-----FSKTESASII 356		
QY	403	GKYMGSGTFGVLTMPR--YNNPPDVNGDLADVYMYGATRPSMT--BPYFAKMRPPGSES 459		
DB	357	APIPISNGDNILVSSQSYNPPFCVN-----FGTDRSSGTSYNDFTNATVFLNGRS 407		
QY	460	HOANVNGFAOITPIKGLTLKAAQAGVDITNRTSSKRMPPNPYDST--PLGERREBAYRDVS 518		
DB	408	YQYNTYNF-QISP-----GLKGRFG-----DSSWQ-----WDATFYGVKVKQS---IN 447		
QY	519	KSFNTAEYKFSIDEXHDLTALMGHEYIEYGDV-----IGASSKGF-----ESDKL 565		
DB	448	NGFLDYATFNOAI-----GPSFLDSGVVKCGSAGAAIAGCTPLNFFNLDSNLL 497		
QY	566	MLLSQGTGNSLSLPEHRVAYEAYLSFPFSRFTYGFDKMWIDFSVRNDQSSRFGSNNRSA 625		
DB	498	ATL-QGMVVNPVITSVTV-----KQPEANANGT 525		
QY	626	WFYSVGMGFDIYNKFIQESNWLSDLRLKMSYGTTCNSEI GYNHQAOLVTNNYTEDAMGL 685		
DB	526	LFDMPAG-----AASLAAGLSYRKERTSTTSDPLMTGDEBDGCGVIEYCASTL-- 573		
QY	686	SISTAGNPDLSEWKQSQFNFGLAAGAFNNRLSAEVDFFVTRTNDMLIDVMPYISGF--- 742		
DB	574	-----SGSFDVK-----EAYAEALPFLIKD--LPGINALNIT 603		
QY	743	-----FSQYQNVGSMKNTGVDLS-----LKGTIYQNKQNVVYASANFYNNROEITKLF 790		

```

Query Match      3.9%; Score 207; DB 16; Length 696;
Best Local Similarity 20.1%; Pred. No. 0.0092;
Matches 178; Conservative 113; Mismatches 280; Indels 316; Gaps 48

Qy 101 LEQVVVLGYGTGCKLSTVSSVAKVSSEKLAEPVIMDALQGVAGQMVTWTSQDPTA 160
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 25 LDSSIVASGFTQDIKEAPATINVTIKELQSKPYRDVAEAI-ADIPGVDLYASKG--- 80
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 161 VASVEIHGTSL-----GASSAPLYIVDGMQTSLDVWATMNPDP-----E 201
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 81 -----TGSYNTMKGITGYTLLVLDGRQGIG--GEVPGFNGEINSGFLPPISSIE 130
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 202 SMSVYLKASATSIYGAAAGWVFIQTKKGWSEGRITFNASVYGISOILATKPLDNWMT 261
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 131 RIEVIKGMST-LYGSALGVVNIITK--KVSK-----WETSVSLDALIN----- 174
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 262 GDELLDFQVAGFWGNQTQVKQXMDILAGAEILYGNVDSLKDBYGKTLFFVDFDNHDADW 321
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 175 --ENKD-----WGNTY-----GTSIYSGPLMVDKLGTLT----- 202
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 322 LKALFKTAPTSQGDISFGSGSQTSYVASIGYFPOEGMARBPANPKRYSGRLNPESSRINE 381
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 203 ---RFREFFYRQOSNVNEFTNGS-----GQRVQGDQAOSPTKANNFN----- 239
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Qy 382 WLKVGANLGAIANRRGADYF-----GKYVMGSGTGVLTMPRYNPPDVNGDLADVY 435
Db 240 ---IGTRIS-YLANDYTFIDIDFSRHYDNKQQLGTITSP-----278
Qy 436 MYGATRESMTPEYPAKMRPFSSHQANVNGFAQITPI-KGLTLKAQAGVDITNTRTSS- 493
Db 279 --GRTPGSLT-----GGYADIMEVDKFTVYLSHEGVVENFSTISGL 317
Qy 494 --KMPNPNPD-----STP--LGERRERAYRDV-----SKSFTNTABYKFSIDEKHDLTALMG 542
Db 318 QYNKVSNDGREVWGQSTQPELGENRDIVAEDIIILDTKSV-----IPLQGSHLS--VG 368
Qy 543 HEY-IEYEGDVIGASSGPFSDKMLLSQKGTGNSLSLPEHRAEYAYLSFFGRFNYGPD 601
Db 369 GEYLEKQKQKI-ASPTNFOYLLAIPAE-----DEYSIKDDLRLTFGARYNH--- 415
Qy 602 KMWYIDFSVRNDQSSRFGSNNRSAMFYSVCGMFDIYNKFIQESNWLSDLRLKMSYGTGN 661
Db 416 -----HEIFG-NNVSPRAYV-----YNP-----TNELTLKGGVSTGPR 448
Qy 662 SEIGNYHQALVTYNNYTEDAMGLSISTAGNPDLSEKQSFNGLAAGAFNNRLSAEVD 721
Db 449 TPYAN-----RLIN-GTYSYSGQG-RPPTYGNPDL--KEETSLNVEIAA-IYNNDL----- 494
Qy 722 FYVRTTNDMLIDVMPYISGFFSOYQNVGSMK--NTGVDLSLKGTYIYONKDNVYASANF 779
Db 495 FYVSAT-----GFLTNPKDKISQSYNNSEPIPGIGTCADR-----CSRAI 536
Qy 780 NYNRQETIKLFFGLNKYMLPNTGTIWEIGYPNSFYMAEYAGIDKKTKQLWYV--PQQVD 837
Db 537 NHGKVEYKGVELGAGISPLDNL-----NNFAYTYLDTVEVKEAQRDRSVIGKPEQDS 587
Qy 838 ADGNKVTTSQYSADLETRIDKSVTPITGSGFSGASWKGSLDLADPAYIVGKWNNDRY 897
Db 588 LKHINMLKTEYS-----FYNKITP-----WIKGEWQI--DRY 617
Qy 898 -----FTENAGGLMOLNK-----DKMLINAW 918
Db 618 MGDNTNINREYKIDIFLASMVGVRDINKQWSINAAIYNLFDKSFTNGW 664

```

Search completed: January 7, 2004, 19:03:13
Job time : 59 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2004, 18:52:09 ; Search time 22 Seconds
(without alignments)
1955.916 Million cell updates/sec

Title: US-09-581-286A-424
Perfect score: 5298
Sequence: 1 MKRMTLFFLLCLLTSGWAMA.....VGNQYPNSKQYVAGIQLSF 1017

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA*
1: /cgn2_6/ptodata/1/1aa/5A-COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B-COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A-COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B-COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCTUS-COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	225.5	4.3	617	US-09-252-991A-30921	Sequence 30921, A
2	201	3.8	2057	US-09-499-203-2	Sequence 2, Appl
3	197	3.7	1536	US-09-206-942-67	Sequence 67, Appl
4	191	3.6	1536	US-08-038-682-2	Sequence 2, Appl
5	191	3.6	1536	US-08-302-832-2	Sequence 2, Appl
6	191	3.6	1536	US-08-530-198-2	Sequence 2, Appl
7	191	3.6	1536	US-08-469-880-2	Sequence 2, Appl
8	191	3.6	1536	US-08-728-470-2	Sequence 2, Appl
9	191	3.6	1536	US-08-617-697-2	Sequence 2, Appl
10	191	3.6	1536	US-08-719-641-2	Sequence 2, Appl
11	189	3.6	663	US-08-765-081-7	Sequence 7, Appl
12	189	3.6	663	US-09-098-082-7	Sequence 7, Appl
13	178	3.4	703	PCT-US95-06994-9	Sequence 9, Appl
14	175.5	3.3	759	US-09-328-352-7650	Sequence 7650, A
15	174	3.3	756	US-09-252-991A-30809	Sequence 30809, A
16	173.5	3.3	1600	US-08-617-697-10	Sequence 10, Appl
17	172.5	3.3	1529	US-08-728-470-10	Sequence 10, Appl
18	172.5	3.3	1529	US-08-719-641-10	Sequence 10, Appl
19	172.5	3.3	2123	US-08-968-685A-10	Sequence 10, Appl
20	168	3.2	1338	US-08-728-470-9	Sequence 9, Appl
21	168	3.2	1338	US-08-719-641-9	Sequence 9, Appl
22	168	3.2	1599	US-08-617-697-9	Sequence 9, Appl
23	166	3.1	761	US-09-328-352-5942	Sequence 5942, Ap
24	165	3.1	1619	US-09-328-352-7347	Sequence 7347, Ap
25	163.5	3.1	643	US-09-328-352-5146	Sequence 5146, Ap
26	162.5	3.1	1346	US-08-471-033-23	Sequence 23, Appl
27	162.5	3.1	1346	US-08-471-044-23	Sequence 23, Appl

28	162.5	3.1	1346	2	US-08-463-483A-23	Sequence 23, Appl
29	162.5	3.1	1346	2	US-08-471-046B-23	Sequence 23, Appl
30	162.5	3.1	1346	2	US-08-470-566B-23	Sequence 23, Appl
31	162.5	3.1	1346	2	US-08-469-334-23	Sequence 23, Appl
32	162.5	3.1	1346	3	US-09-300-529-23	Sequence 23, Appl
33	160.5	3.0	703	5	PCT-US95-06994-8	Sequence 8, Appl
34	160	3.0	652	1	US-08-765-081-6	Sequence 6, Appl
35	160	3.0	652	3	US-09-098-082-6	Sequence 6, Appl
36	160	3.0	703	4	US-09-328-352-7467	Sequence 7467, Ap
37	160	3.0	775	4	US-09-328-352-7095	Sequence 7095, Ap
38	160	3.0	1338	1	US-08-471-033-50	Sequence 50, Appl
39	160	3.0	1338	2	US-08-471-044-50	Sequence 50, Appl
40	160	3.0	1338	2	US-08-463-483A-50	Sequence 50, Appl
41	160	3.0	1338	2	US-08-471-046A-50	Sequence 50, Appl
42	160	3.0	1338	2	US-08-470-566B-50	Sequence 50, Appl
43	160	3.0	1338	2	US-08-469-334-50	Sequence 50, Appl
44	160	3.0	1338	3	US-09-300-529-50	Sequence 50, Appl
45	160	3.0	1612	1	US-08-169-927-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-09-252-991A-30921
; Sequence 30921, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30921
; LENGTH: 617
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30921

Query Match	4.3%	Score 225.5;	DB 4;	Length 617;
Best Local Similarity	18.7%	Fred. No. 5.8e-10;		
Matches 157;	Conservative 105;	Mismatches 265;	Indels 313;	Gaps 32;
QY	79	MTTKEVALANYMKIVLDPDSKVLQVVLVGGTGTGQKLTSGSVAKVSSEKLAERKPVANI	138	
Db	7	LTPRAVALCGASSLSLAEPVSLAQVVT-ATRTAQATASQSLAANSVIDREDI-ERSQARS	64	
QY	139	MDALQGVAGQVMMTSGDPTAVASVEITHGTSLGASSAPLYIVDMQ-----TSLDVVA	193	
Db	65	VPPELLRQVPGVS-LANNGGFKNTTLPLRGT-----ESDHLVLVIDGIVKVSASAGLTAQ	119	
QY	194	TWNPNDPESMVLKDDASATSIYGARAANGVVFIQTKKGMSEGRITFNASIGISQILNT	253	
Db	120	DLPVELIERIEVVR-GPRSSLYGSEAGVGIQIFTRRGD-CQGAKPFFSAGYGTHTQ----	173	
QY	254	KPLDNMTGDELLDFQVKAGFMGNQTVQVKVMDLAGAEDLYGNDYSLDKBYGKTLFPV	313	
Db	174	-----	173	
QY	314	DFNHADWLKALFKTAPTSGQDISFSGSQGTSYVASIGYFQDEGMAREPA-----	364	
Db	174	-----TLEGSAGVSGAGNGWYSLGVSSFDTAGINTKRAGTAGYEPR	216	
QY	365	-NPKRYSGRLNPFGRINELWLVKANLSCAIANRRSADYFGKYMGSGTGFVLTPRYNP	423	
Db	217	DGYRNLGNLRGGYRFDN-----GLELDGTLRAKSHNDYDQVFGNSG-----	259	

QY 424 FVNGDLADVYMYGATRPSTMTPEYFAKMRPFSESHQANVNGFAQITPIKGLTLKAQAG 483
Db FANAD-----GEQNLVGGARFTPDWLVTLQAG 290
QY 484 VDINTRTSSKRMNNPYDSTPLGERRERAYRDVSKSFTNTAEYKFSIDEXHDLTALMGH 543
Db -----RSEDKADAYQD--GRFYSRFDTRRDSLSWQNDLTLAEGH 327
QY 544 EY---IEYEGDVIGASSKGFESDKMLLSOGKTGNSLSLPEHRVAEYAYLSFFSRF--NY 598
Db VLTIGYDQKDEI--SSSEAFSVDS--RLNK-----WFAQYLGQY 364
QY 599 GFDKMYIDFSVRNDQSSRFQ--SNRESAFYSGMGFDIYNKFIQESNMLSD--LRLKMS 655
Db GRQDM---QLSLRDDNQGVHDTGSAWGYA-----LSDALRFTVN 404
QY 656 YGTTGSEIGNYHQAALVNNYTTEDAMGLSISAGNPDLSWEKQSFNGL 707
Db YGTAFKAPTEN-----ELYYPDYGNDPLDAETSRSELEVGLSGTHGWGH 447
QY 708 -AAGAFNNRLSAEVDYFVRTNDMLIDVPMPIYISGFESQYQNVGSMKNTGVDLSLKTIIY 766
Db WAVNAFTNV-----DDLIGNDPRPAPGRPGQPNNDIARIRGVELVL--GSQW 495
QY 767 QNKDNNVYASANFNRYQ-----EITKLPFGLN-----KYM 797
Db LQWNNANATFLDPQNRSGVNDGNEPLPRARPMFNLEDRRERLSGASVHAEGRYD 555
QY 798 LPNTGTIWEIGYPSFYMAEYAGIDKKTGKQL---WYVPGQVDADGNKVTTSQYSADLET 854
Db -----ANKVRLGGVATDLRSEYRLNDEWLQGR-----ANLFGADYET 597

RESULT 2

US-09-499-203-2
; Sequence 2, Application US/09499203
; Patent No. 6570065
; GENERAL INFORMATION:
; APPLICANT: KOSSMANN, Jens
; APPLICANT: WELSH, Thomas
; APPLICANT: QUANZ, Martin
; APPLICANT: KNUTH, Karola
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase
; FILE REFERENCE: 147-196P
; CURRENT APPLICATION NUMBER: US/09/499,203
; CURRENT FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2057
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-499-203-2

Query Match 3.8%; Score 201; DB 4; Length 2057;
Best Local Similarity 19.4%; Pred. No. 4.7e-07;
Matches 232; Conservative 159; Mismatches 414; Indels 388; Gaps 59;

QY 27 GTVISSEBNEPLIGANVVVVGNTTIGAAATLDGNTLSVSPANAKMLRVSYSGMTTKEVAI 86
Db GTNDSGEKKVPVPTNDSLSKQGTGDFWYSDGN-----RVD----- 90
QY 87 ANVMKIVLDPDSKVLEQVVLGYGTQKLSVSSSAKVSSEKLAEK--PVAN--IMDALQ 143
Db -----QKTNQILLTAEBQLKKKNEKNLSVISDTSKDDENISKQTKIANQQTVDATAK 142
QY 144 GOVAGMQVMTTS--GDPTAVASVEIHG-----TGSLSGASAPLYIVDGMQTSLDVVATMN 196
Db -----LTTNSLSDITGTHYENHNGYFYVIDASQVTKQVILQNIQNLQYFD----- 189
QY 197 PNFESMVLKDSATSIY-----GARAANGVVFQTK-----KGRMSERGRITFNASYG 246
Db DNGYQVKGSRFDVNGKHIIYDPSVTGKASSNVDIVNGKAQGYDAQGNQLKKSYVADSSGQT 249

QY 247 ISOILNTKPLDNMTGDELLDF-----QVKAGFMG--NNOTVQVKVDM--ILAGABDLYG 297
Db -----YFPDNGQPLIGLQITDGNLQYFNQGVQIKGGFDVNNKRIYPAPNTGNVANTELING 309
QY 298 -----NYDSLKDEYKGTLPFPVDNHDADMLKALFKTAPTSGQDISFSGSGSGTSYIAS 350
Db KLQGRDANGNQVKNAPSKDVAGNTFYFDANGVNL-----TGLQITIS-----GKTY--- 354
QY 351 IGYFDEGEMAREPANFKRYSGRNLNPFESRINEMLKVCANLSG-----AIANR 396
Db -----YLDQGHRL-----KNYAGTFN-----NQPMYFADTGAGKTAIEYQFDQGLVSQNE 402
QY 397 RSADYEGKYMGs---GTFGLVTPRYNPFV--NGDLADVYMYGATRPSTMTPEYFAK 451
Db NTHPNAKSYDKSKSFENVGYLTADTWYRPTDILKNGD-----TWASTE---TD 449
QY 452 MRP-----FSSESHQANVNGFAQITPIKGLTLKAQAGVDITNTRTSSKRMNN----- 499
Db MRPLMTWMPDKQTQANLNFMS---SKGL-----GITTTVTAATSQKTLNDAAFVQIT 500
QY 500 PYDSTPLGERRERAYRDVSKSFTNTAE--YKFSIDEXHDLTALMGHEYIEYEGDVIGASS 557
Db AIBQQISLKKSTEWLRDAIDSFVKTOANWNKQTEAFDGLQWLOGGFLAYODD-----SH 556
QY 558 KGFESDKMLLSOGKTGNSLSLPEHRVAEYAYLSFFSRFNGYFDPKMWYIDFSVRNDQSSR 617
Db -----SGNNRKLGRPI-----NIDGS---KUTTDG 585
QY 618 FGSNNRSAFYSGMGFDIYNKFIQ--ESNWLSDLRLKMSYGT--TGNSEIGNYHQAALVT 674
Db KGSE-----FLLANDIDNSNPIVQAEQLNWLHYL---MNFGSIITGNNDNANFDIRVDA 636
QY 675 VNYTEDAMGLS-----ISTAGNPDLs--WE 698
Db VDNVDADLLKIAGDYFKALYGTDKSDANANKHLSILEDWNGKDQPVYVNOQGNLAQLTMDYT 696
QY 699 KGSQNFGLAAGAFNNRLS---AEVDYFVR--TTNDMLIDVPMPIYISGFESQYQNVGSMK 753
Db VTSQFQNSLTHGA--NNRSNNWYFLDTGYLYNGDLNKKIVDKRPNSTGLVNRITANGSDTK 755
QY 754 -----NTGVDLSLKGTTIYQNKDMNVYASANFNRYNRQBITKL 789
Db VPIYNSFVRAHDYDAQDPIRKAMIDHGIKMQDPTFTDQ---LAQMEFYFKDQENPSG 812
QY 790 FGLNKYMLPNTGTIWEIGYPSFYMAEYAGI--DKTKGQLWYVPGQVDADGNK----- 842
Db FKKYNDYNLFSA-----YAMLLTNKDTVPVYV--GDMYLEGGQYMEKG 854
QY 843 -VTTQSYSADLETRIDKSVTPPTITGGFSLGASKWGLSL---DADPAYIV--GKWMINNDR 896
Db TIYNFVISALLKARIKY-----VSGQGTWATDSSGKDLKQDGTDLTTSVRFKGITSDQ 909
QY 897 YFTE-----NAG-GL-----MOLNKDKML-----LNAW 918
Db -----YKGF-----DPEAGNV 998
QY 910 TTTQDNSQYKQKQIGVIVGNPNPDLKLANDKTTILHMGAKHKNQLYRALVLSNDSIDVY 969
QY 919 TEDNKETDVPKQSQSPQFDTHLLENASFLRLKMLKLTYYVLPNSL-----FAGQNVIGGA 972
Db DSDDK---APTARTNDNGDLIFHKITNTFVKQDGTIINYEKMGSLNALISGYLGLVWVPVGA 1026
QY 973 RYVLMARNLLTVTK-----YKGF-----DPEAGNV 998
Db SDSQDARTVATESSSSSNDGVSFHSNALDSNVIEGFSNFQAMPTSPQSTNV 1079

RESULT 3

US-09-206-942-67
; Sequence 67, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping

APPLICANT: Klein, Michel H.
TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
FILE REFERENCE: Molecular Weight Proteins
CURRENT APPLICATION NUMBER: US/09/206,942
CURRENT FILING DATE: 1998-12-08
EARLIER APPLICATION NUMBER: 09/167,568
NUMBER OF SEQ ID NOS: 95
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 67
LENGTH: 1536
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-09-206-942-67

Query Match 3.7%; Score 197; DB 4; Length 1536;
Best Local Similarity 18.7%; Pred. No. 6.2e-07;
Matches 206; Conservative 151; Mismatches 377; Indels 368; Gaps 51;

QY 25 VKGTVISSEDEPLIGANVVVG-----NTTIGAAATDLDGNFTLSVPANAKMLRVSYSGM 79
DB 136 LKG-ILDSNGQVFLNPNGITIGKDAINTNGFTASTLD-----ISNENIKARNFYFE 187

QY 80 TKEVAIANVMK---IVLDPD-----SKVLEQVVLGYGTGQKLSVSGSVAKVSSEK 129
DB 188 QTKKALAEIVNHGLITVKGDSVNLIGGKVNKEGVI-----SVNGGSISLLAGQK 238

QY 130 LAEKPVAN-----IMDALQGVAGQVMTTSGDPTAVASVEIHGTGSLGASSAPLYVDG 184
DB 239 ITISDIINPTIYSAAPENEAIVGLDIPAKGGINVRAATIRNOGKLSAD-----289

QY 185 MOTSLDVVATMNPDPFESVLKDSATSIYGARAA---NGVWFIQTKKQMSRGRIT 240
DB 290 -----SVSKDKSGNIVLSAKGEAIEIGVISAQNOQAK---GGKL- 326

QY 241 FNASGISOILNKPDLNMTGDE-----LLDFQVKAG---FWGNQTVQVKDMILA 290
DB 327 -----MITGDKVTLTKTGAVIDLSGEGGETYLGDERGEGKGIQIA 368

QY 291 GAEDLYGNVD---SLKDEYKTLF-----PVDFNHDAWLKALFKTAPTSQGDIFSQGS 342
DB 369 KITSLEKSGTINVSKEGKGRAIVMGDIALIDGNINAQ-----GSGDIAKTGTF 417

QY 343 QGTS-----YYASIGYFDQSGMAREPANFK---RYSGLNPFESRINEWLKVGANLSCATAN 395
DB 418 VETSGHDLFIKNAIVDAKEMLLDPDVSINAEAGRSN-TSEDDYETSGNSASTPKRN 476

QY 396 RRSADYFG---KYMGSGTFFGLVTPRYNPPDVNGDLADVYMYGATRPSMTPEYPAKM 452
DB 477 KEKTLTNTTLESILKKGTFVNIT-----ANQ 503

QY 453 RPFSESHQANVNGFAQITPIKGLTL---KAAQAGVDINTTSSKMPNPNYDSTPLGE 508
DB 504 RIYVNSSINLS-NG-----SLTWSRGGGGVGEINNDITT-----GD 540

QY 509 RRERA-----YRDVKSFTNTAYKFSIDEKDLTALMG-HEVIEYEGDVIGASSKG 559
DB 541 DTRGANLITYSGWDVHKNSILGAGQNNITAKQDIAFEKSGNQVITCGGHTITSGNQKG 600

QY 560 FESDKMLLSQKGTGNSLSLPEHRAVAYLSPFRPNYGFDMWYIDFSVRNDQSSRFG 619
DB 601 FRPNNVSL---NGTSGGLQFTTKRTNKYAITNKE-----G 633

QY 620 SNRGAFWYVSGMFDI-----YKFTQESNW-LSDLRLKMSYGTGNGSEIGNV 668
DB 634 TLN-----ISGVNISMVLPKNSGYDKFGRTYWNLTSLNV-----SESGGFN 677

QY 669 HQALVTNNVYTDAMGLSISTAGNPDLSWEKOSQNFGLAAGFNNRLSAEVDYVYRTTN 728
DB 678 -----LTIDSRGSDSAGTLTQPNLNGISFNKDTTFNV-----ERNARVNF-----718

QY 729 DMLIDVMP-----YISGFFSQYQ-NVSGMKNTGVDSL-----KGTIYQNKDMNV 773

DB 719 ----DIKAPIGINKYSSLNYSFNGNISVSGGSDFTLLASSSNVQTPGVVINSKYPNV 774
QY 774 YASANFNVRQBITKLPFLKNKMYLPN-TGTIWEIGYFNSFMAEYAGIDKTKGKQLWTV 832
DB 775 STGSSLRKFTSGSTKTSFSEIKDLTLNATG-----GNITLQVEGTGDMIGKI-VA 825

QY 833 PQQVDADGNKVTTQYSADLETRIDKSVT-----PPITGGFSLGASW 874
DB 826 KKNITFEGGNITFGSRKA--VTEIEGNTVINNNANVTILGSDFDNHQKPLTIKQDVINS 883

QY 875 KGLSLDADPAFVIVGKWMIND-----RYFTENAGGLMQLNKDKMLNANTWEDNKETDVP 928
DB 884 GNLTAGGNIVNIAGNLTVESNANFKAITNFTFNVGGLFD-----NKGNSNISIA 932

QY 929 KLGQSPQDPTLLENASPLRLKLNKLTVVLNSLFPAGONVIGGARVYLWARNL-----L 982
DB 933 KGG-----ARFKDIDNSKNLSITNSSSTYRTIISG-----NITKNGDL 972

QY 983 TVTRYKGFDPPE--AGGNVGNKQ 1002
DB 973 NITN-EGSDTEMOIGDVSQKE 993

RESULT 4
US-08-038-682-2
Sequence 2, Application US/08038682
Patent No. 5549897
GENERAL INFORMATION:
APPLICANT: BARENKAMP, STEPHEN J
APPLICANT: ST. GEME III, JOSEPH W
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/038,682
FILING DATE: 16-MAR-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BERKSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-293
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-038-682-2

Query Match 3.6%; Score 191; DB 1; Length 1536;
Best Local Similarity 19.1%; Pred. No. 2e-06;
Matches 210; Conservative 146; Mismatches 381; Indels 362; Gaps 54;
QY 25 VKGTVISSEDEPLIGANVVVG-----NTTIGAAATDLDGNFTLSVPANAKMLRVSYSGM 79

```
Db 136 LKG-ILDSNGQVFLINPNGITIGKDAIINTNGFTASTLD-----ISNENIKARNPTFE 187
Qy 80 TTKEVAIANVMK---IVLDPD-----SKVLEQVVVLGYGTGQKLTSTVSGSVAKUSSEK 129
Db 188 QTKDKALABIVNHGLITVGDGGSVNLIGGKVKNEGVI-----SVNGGSISLLAGQK 238
Qy 130 LAEPFVAN-----IMDALQGVAGQVMTTSGDPTAVASVEIHGTSIGLASSAPLYIVDG 184
Db 239 ITISDIINPTITYSIAAPENAVNLGDI FAKGGINVRAATIRNOGKLSAD----- 289
Qy 185 MQTSLDVAATNPNDPESMVLKDSATSIYGARAA---NGVVFIQTKKGKMSERGIT 240
Db 290 -----SVSKDKSGNIVLSAKEGEAEIGGVI SAQNOQAK---CGKL- 326
Qy 241 FNASYGISQILNTKPLDNMMTGDE-----LLDFQVKAG---FWGNNTQVQVKDMILA 290
Db 327 -----MITGDKVTLTKTGAVIDLSGKEGGETYLGDERGEGKNGIQLA 368
Qy 291 GAEDLYGNYD---SLKDEYGTLPFPVDFNHDADWLKALFKTAPTSGQDISFSGG---SQG 344
Db 369 KXTSLEKSGTINVSKEKGGRAIV-----W-----GDIALIDGNINAQG 407
Qy 345 TSYVASIYGFDOEGMAREPANFKYSGRLNF--ESRI---NEWL-----KYCANLSGAIAN 395
Db 408 SGDIKATG-----GFVETSGHDLFIKDNAIVDAKEWLLDFDNVSNINAEATAGRS 455
Qy 396 RRSADYFGKYMGSGTGVLTMPRYNPPFVNGDLADVYMYGNATRPSMTPEPAKWRPF 455
Db 456 NTSBD---DEYTGSGNSASTPKRKNKEKTLTNTLTLESII--LKKGTFFVNIT---ANQRIY 506
Qy 456 SSES HQANVGFAQITPIKGLTL---KAAQAGVDITNTRTSSKRMPPNPDSTPLGERRE 511
Db 507 VNSSINLS-NG-----SLTLWSEGRSGGVEINNDIIT-----GDDTR 543
Qy 512 RA-----YRDVSKFTNTAETKFSIDEKHDLTALMG-HEYIEYEGDVIGASKGFES 562
Db 544 GANLTIYSGGWDVHKNSILSQAQGNINITAKODIAPEKSGNOVITGGTITSGNKGFRF 603
Qy 563 DKMLLSQGTGNSLSLPEHRVABAYLSPFRSRYNGFDKWMYIDFSVRNDQSRFGSNN 622
Db 604 NNVSLL---NGTSGSLQPTTKRTKNYAJTNRKFE-----GTLN 636
Qy 623 RSAMFYSGVGMPDI-----YNKFIQESNW-LSDLRLKMSYGTGTGNSIENYVHQA 671
Db 637 -----ISGKVNLSMVLPNKESGYDKFKGTYWNLTSLVN-----SESGEYN--- 677
Qy 672 LVTNNYTEDAMGLSISTAGNPLDSWEKQSQFNPLAAGAPNNRLSAEVDFFYVRTNDML 731
Db 678 -LTIDSRGSDAGTLTQPNYNLGISFNKDTTFNV-----ERNARVNF----- 718
Qy 732 IDVPMP-----YISGFFSQYQ-NVGSMMKNTGVDLSL-----KGTIYQKDMNYYAS 776
Db 719 -DIKAPITGINKYSLNAYSFNGNISVSOGGSVDPTLLASSNNVQTPGVWINSKYFNVTG 777
Qy 777 ANFNYNROETIKLFFGLNKLWMLPN-TGTIWEIGYPNSFYMAEYAGIDKTKGKOLWYVPGQ 835
Db 778 SSURFKTSGSTKGFSEKDLTLNATG-----GNITLQVEGTDGMLKGKI-VAKKN 828
Qy 836 VDADGNKVTTSQYSADLETRIDKSVT-----PPITGGFSLSGASWKGL 877
Db 829 ITFEGGNITFGSKRA--VTEIEGNVTINNANVTLIGSDPDNHQKPLTIKKDVIINSGLN 886
Qy 878 SLADPAIYVVKMINND-----RYFTENAGGLMQLNKKMLLNWATEDNKETDVPKLG 931
Db 887 TAGGINIYIAGNLTVESNANFKAITNFTFNVGGLFD-----NKGNSNISIAKGG 935
Qy 932 QSPQDFTHLLENASFLRKLKLLTYLPLNSLFAQONVIGGARVYLMARNL-----LTVT 985
Db 936 -----ARFKDIDSKNLSITNSSTVTTIISG-----NITNKGNDLNIIT 975
Qy 986 KYGFDPE--AGGNVGNQK 1002
Db 976 N-EGSDTEMQIGGDVDSQKE 993
```

RESULT 5

```
US-08-302-832-2
; Sequence 2, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: Of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-302-832-2
```

Query Match 3.6%; Score 191; DB 1; Length 1536;

Best Local Similarity 19.1%; Pred. No. 2e-06;

Matches 210; Conservative 146; Mismatches 381; Indels 362; Gaps 54;

Qy 25 VKGTIVSSDNBPLICANVVVVG-----NTTIGAATDLGDNFTLSVPANAKMLRVSYSGM 79

Db 136 LKG-ILDSNGQVFLINPNGITIGKDAIINTNGFTASTLD-----ISNENIKARNPTFE 187

Qy 80 TTKEVAIANVMK---IVLDPD-----SKVLEQVVVLGYGTGQKLTSTVSGSVAKUSSEK 129

Db 188 QTKDKALABIVNHGLITVGDGGSVNLIGGKVKNEGVI-----SVNGGSISLLAGQK 238

Qy 130 LAEPFVAN-----IMDALQGVAGQVMTTSGDPTAVASVEIHGTSIGLASSAPLYIVDG 184

Db 239 ITISDIINPTITYSIAAPENAVNLGDI FAKGGINVRAATIRNOGKLSAD----- 289

Qy 185 MQTSLDVAATNPNDPESMVLKDSATSIYGARAA---NGVVFIQTKKGKMSERGIT 240

Db 290 -----SVSKDKSGNIVLSAKEGEAEIGGVI SAQNOQAK---CGKL- 326

Qy 241 FNASYGISQILNTKPLDNMMTGDE-----LLDFQVKAG---FWGNNTQVQVKDMILA 290

Db 327 -----MITGDKVTLTKTGAVIDLSGKEGGETYLGDERGEGKNGIQLA 368

```

QY 291 GAEDLYGNVD---SLKDEYKTLFPVDFNHDADWLKALFKTAPTSGQDISFSGG---SQG 344
Db 369 KKTSLKSGSTINVSQKGGRAIV-----W-----GDIALIDGNAQ 407
QY 345 TSYVASIGYFDOEGMAREPANFKYSGRLNF--ESRI---NEWL---KVGANLSGATAN 395
Db 408 SCDIAKTG-----GFVETSGHDLFTKDAIVDAKEWLLDFDNVSNATAGRS 455
QY 396 RRSADYFGKYMGSGTGLTTPRYNPFVNGDLADVYMYGATRPSTPEYFAKMPFF 455
Db 456 NTSED---DEYTGSGNSASTPRKNEKTLTNTTLES1--LKKGTFFVNT---ANQRIY 506
QY 456 SSESQANVNGFAQITPIKGLTL-----KAQAGVDITNTRTSKRMNPNPDSTPLGERRE 511
Db 507 VNSSINLS-NG-----SLTLWSEGRSGGVEINNDITT-----GDDTR 543
QY 512 RA-----YRDVSKSFNTAEYKFSIDEXHDLTALMG-HEYIEYEGDVIGASSKGPES 562
Db 544 GANLTIYGGWVDVHKNSISLGAQGNINITAKQDIAFEKGSNQVITGQGTITSNGKGRPF 603
QY 563 DKMLLSQCKTGNLSLPHRVAEYAYLSFFSRFNYGDKWYIDFSVRNDQSSRFSGNN 622
Db 604 NNVSLL---NGTSGGLQFTTKRKNAITNKFE-----GTLN 636
QY 623 RSMFYVSGMPDI-----YNKFIQESNW-LSDLRLKMSYGTGTGNSIGNYHQA 671
Db 637 -----ISGKVNISMVLPKNSGYDKPKGRTYWNLTSLNV-----SESGEFN--- 677
QY 672 LTVNANYTEDANGLISITAGNPDLSWESQSFNFGLAGAFNNRLSAEVDVFRNTDML 731
Db 678 -LTIIDSRGSDSAGTUTQPYNLNGISFNKDTTFNV-----ERNARVNF----- 718
QY 732 IDVPMP-----YISGFSSOYO-NVGSKMNTGVDLSL-----KGTIYQNKDNVYAS 776
Db 719 -DIKAPIGINKYSSLYASFNGNISVSGGSDVDFILLASSNVQFVGVINSKYFNVTG 777
QY 777 ANFNYNROBITKFLGLNKYMLPN-TGTIWEIGYPNSFYMAEYAGIDKTKGQLWYFQG 835
Db 778 SSLREKTSSTKTFGSIEKDLTLNATG-----GNITLLQVEGTGDMIGKI-VAKKN 828
QY 836 VDADGNKVTTSQYADLETRDKSVT-----PPYTGFSGLASWGL 877
Db 829 ITFEGNITFGSRKA--VTEIEGNTVINANNVTILIGSDFDNHQKPLTIKKDVIINSGL 886
QY 878 SLDAFAYIVGKWMINND-----RYFTENAGGLQNLKDKMLNATWEDNKETDVPKLG 931
Db 887 TAGGNIVNAGNLTVESNANFKAITNFTNVGGLFD-----NKGNSNISIAGGG 935
QY 932 QSPQDTHLLENASFLRLKNLKLYVLNPSLFAGONVIGGARVYLMARNL-----LATV 985
Db 936 -----ARFKDIDNSKNLSITTSSTYRTIISG-----NITNKGDLNIT 975
QY 986 KYKGFDPB--ACGNVKGKQ 1002
Db 976 N-EGSDTEMQIGGDVSQKE 993

```

RESULT 6

US-08-530-198-2

```

; Sequence 2, Application US/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington

```

```

; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,198
; FILING DATE: 13-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: JWB-1186
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-530-198-2

```

Query Match 3.6%; Score 191; DB 2; Length 1536;

Best Local Similarity 19.1%; Pred. No. 2e-06;
Matches 210; Conservative 146; Mismatches 381; Indels 362; Gaps 54;

```

QY 25 VKGTVISSEDEPLIGANVVVVG-----NTTIGATDLGDGNTLSVSPANAKMLRVSYSGM 79
Db 136 LKG-ILDSNGQVFLNPNGIIGKDAIINTNGFTASTLD-----ISNEIKARNFTPE 187
QY 80 TTKEVAIANVMK---IVLDDP-----SKVLBQVVVLGYGTGQKLSVSGSVAKVSBEK 129
Db 188 QTKKALAEIVNHGLITVGDGVSNNLIGGKVKNEGVI-----SVNGGSISLAGQK 238
QY 130 LAEPVAN-----TMDALQGVAGVQVMTTSGDPTAVASVEIHGTGSLGASSAPLYVDG 184
Db 239 ITISDIINPTITYSIAAPENEAVALGDIFAKGGINVRAATIRNQGKLSAD----- 289
QY 185 MOTSLDVVATWNPDPFESMVLKDAATSIVGARAA-----NGWVFIQTKGQMSERGRIT 240
Db 230 -----SVSKDGSNIVLSAKEGEAEIGSVISAQOQAK---GGKL- 326
QY 241 FNASYGISQILNTRPLDNMTGDE-----LLDFQVRKAG---FMGNNQTVQKVDMLA 290
Db 327 -----MITGDKVTLKTGAVIDLSGKEGGETYLGGERGEGKNGIQLA 368
QY 291 GAEDLYGNVD---SLKDEYKTLFPVDFNHDADWLKALFKTAPTSGQDISFSGG---SQG 344
Db 369 KKTSLKSGSTINVSQKGGRAIV-----W-----GDIALIDGNAQ 407
QY 345 TSYVASIGYFDOEGMAREPANFKYSGRLNF--ESRI---NEWL---KVGANLSGATAN 395
Db 408 SCDIAKTG-----GFVETSGHDLFTKDAIVDAKEWLLDFDNVSNATAGRS 455
QY 396 RRSADYFGKYMGSGTGLTTPRYNPFVNGDLADVYMYGATRPSTPEYFAKMPFF 455
Db 456 NTSED---DEYTGSGNSASTPRKNEKTLTNTTLES1--LKKGTFFVNT---ANQRIY 506
QY 456 SSESQANVNGFAQITPIKGLTL-----KAQAGVDITNTRTSKRMNPNPDSTPLGERRE 511
Db 507 VNSSINLS-NG-----SLTLWSEGRSGGVEINNDITT-----GDDTR 543
QY 512 RA-----YRDVSKSFNTAEYKFSIDEXHDLTALMG-HEYIEYEGDVIGASSKGPES 562
Db 544 GANLTIYGGWVDVHKNSISLGAQGNINITAKQDIAFEKGSNQVITGQGTITSNGKGRPF 603
QY 563 DKMLLSQCKTGNLSLPHRVAEYAYLSFFSRFNYGDKWYIDFSVRNDQSSRFSGNN 622

```

604 NNVSLL---NGTSGGLQTTTKTKNKYAITNKFE-----GTLN 636
623 RSWFYVSGMFDI-----YNKFIQESNW-LSDLRLKMSYGTGTGNSEIGNYHQA 671
637 -----ISGKVNISMVLPKNESGYDKFGRTYNNLTSLN-----SESGEFN--- 677
672 LVTNNVTEDAMGLSISTAGNPDLSEWESQFNFGLAAGAFNNRLSAEVDYFYVRTNDML 731
678 -LHDSRGSDSAGTLTQPNYLNGLSPNKDTTFNV-----ERNARVNF----- 718
732 IDVMP-----YISGFFSOYO-NVGSMMKNTGVDLSL-----KGTIYQNKDMNYAS 776
719 -DIKAPIGINKYSSLNVSFNGNISVSGGSDVDTLLASSSNVOTPGWINSKYFNVTG 777
777 ANFNYNQETIKLFFGLNKYMLPN-TGTWEIGYFNPFFMAEYAGIDKYGKQLWYVPGQ 835
778 SSLRFKTSGSTGTGFSIEKOLTLNATG-----GNITLQVEGTGDMIGKI-VAKN 828
836 VDADGNKVTTSQYSADLETRIDKSVT-----PRITGGFSLGASWKL 877
829 ITPEGGNITGSRKA--VTELEGNVTINNANVTLIGSDFDNHQKPLTIKDVINSGLN 886
878 SLDAFPAYIVGKMMINND-----RYFTENAGGLMQLNKKMLLNAWTEDNKETDVPKLG 931
887 TAGGNIVNIAGNLTVESNANFKAITNFTFNVGGLFD-----NKGNSNISIAKGG 935
932 QSPQDTHLLNENFLRLKNLKYVLPNSLFAQGNVIGARVYLMARNL-----LTVT 985
936 -----ARFKIDNSKNLSITNSSTVRTIISG-----NITNKGNDLNI 975
986 KYKGFDPE--AGGNVGNQ 1002
976 N-EGSDTEMQIGDVSQKE 993

RESULT 7

US-08-469-880-2
; Sequence 2, Application US/08469880
; Patent No. 5876733
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,880
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstreser, Jerry W

; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-516 MIS:VG
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-469-880-2

Query Match 3.6%; Score 191; DB 2; Length 1536;
Best Local Similarity 19.1%; Pred. No. 2e-06;
Matches 210; Conservative 146; Mismatches 381; Indels 362; Gaps 54;

QY 25 VKGTVISSDNEPLIGANVVVVG-----NTTIGAAATDLDCNFTLSVPANAKMLRVSYSGM 79
DB 136 LKG-ILDSNGQVFLNPNGITIGKDAIINTNGFTASTLD-----ISNENIKARNFTPE 187
QY 80 TTKEVAIANVMK---IVLDPD-----SKVLEQVVVLGYGTGCKLSTVSGSVAKVSSEK 129
DB 188 QTKDKALAEIVNHLITVKGDSVNLIGGKVKNEGVI-----SVNGGSISLLAGQK 238
QY 130 LAEKPVAN-----IMDALQOVAGMQVMTTSGDPTAVASVEIHGTGSLGASSAPLYVDG 184
DB 239 ITISDIINPTITVSIAAPENAVNLGDI FAKGGINVRAATIRNQGKLSAD----- 289
QY 185 MQTSLDVATMNPDPESMSVLKDSATSIYGARAA-----NGVVFITQTKGKMSERGIT 240
DB 290 -----SVSKDKSGNIVLSAKEGEAEIGGVISAQNOQAK---GGKL- 326
QY 241 FNASIGISQILNTKPLDNMMTGDE-----LLDFQVKAG---FWGNQTVQVKOMILA 290
DB 327 -----MITGDKVTLKTGAVIDLSGKEGTYLGDERGEGKNGIOLA 368
QY 291 GAEDLYGNVD---SLKDEYKTLFPVDFNHDAWMLKALFKTAPTSGDISFSFG---SOG 344
DB 369 KKTSLKSGSTINVSKEKGRATV-----W-----GDIALIDGNINAQG 407
QY 345 TSYVASIGYFDOEGMAREPANFKRYSGRNLF--ESRI---NEWL-----KVGANLSGAIA 395
DB 408 SGDIATG-----GFVETSGHDLFIKDNAIVDAKEWLLDFDNDVNSINAEIAGRS 455
QY 396 RRSADYFGKYMGSGTGFVLTPRYNPFVNGDLADVYMYGATRPSMTEPIFAKMRPF 455
DB 456 NTSED---DEYTGSGNSASTPKRNEKKTTLTNTTLESII--LKKGTFTVNIIT---ANQRIY 506
QY 456 SSESQANVNGPAQITPIKGLTL---KAQAGVDITNRTTSSKRMNPNPYDSTPLGERRE 511
DB 507 VNSSINLS-NG-----SLTWSRSGSGGVEINNDITT-----GDDTR 543
QY 512 RA-----YRDVSKSFTNTAEYKFSIDEKHDLTALMG-HEYIEGVDVIGASSKGFS 562
DB 544 GANLTIYSGWVDVHNKISLGAQGNINITAKQDIAPEKSGNOVITGQTTISGKQGRF 603
QY 563 DKMLLSQKGTGNSLSLPEHRAEYAYLGFPSRFNYGFDKWMYIDFSVRNDQSRFGSNN 622
DB 604 NNVSLL---NGTSGGLQTTTKTKNKYAITNKFE-----GTLN 636
QY 623 RSWFYVSGMFDI-----YNKFIQESNW-LSDLRLKMSYGTGTGNSEIGNYHQA 671
DB 637 -----ISGKVNISMVLPKNESGYDKFGRTYNNLTSLN-----SESGEFN--- 677
QY 672 LVTNNVTEDAMGLSISTAGNPDLSEWESQFNFGLAAGAFNNRLSAEVDYFYVRTNDML 731
DB 678 -LHDSRGSDSAGTLTQPNYLNGLSPNKDTTFNV-----ERNARVNF----- 718
QY 732 IDVMP-----YISGFFSOYO-NVGSMMKNTGVDLSL-----KGTIYQNKDMNYAS 776
DB 719 -DIKAPIGINKYSSLNVSFNGNISVSGGSDVDTLLASSSNVOTPGWINSKYFNVTG 777

Db 936 -----ARPKDIDNSKNLSITNTSSSTVYRTIISG-----NITKNGDLNIT 975

Qy 986 KYKGFDP--AGGNVGNQ 1002

Db 976 N-EGSDTEMQIGGDVSQKE 993

RESULT 9

US-08-617-697-2

; Sequence 2, Application US/08617697

; Patent No. 5977336

; GENERAL INFORMATION:

; APPLICANT: Barenkamp, Stephen J

; TITLE OF INVENTION: High Molecular Weight Surface Proteins

; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Shoemaker and Mattare, Ltd.

; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

; STREET: Bldg. 1

; CITY: Arlington

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22202-0286

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/617,697

; FILING DATE: 01-APR-1996

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/302,832

; FILING DATE: 05-OCT-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US PCT/US93/02166

; FILING DATE: 16-MAR-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Berkstresser, Jerry W

; REGISTRATION NUMBER: 22,651

; REFERENCE/DOCKET NUMBER: 1038-557

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 415-0810

; TELEFAX: (703) 415-0813

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1536 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-08-617-697-2

Query Match 3.68; Score 191; DB 2; Length 1536;

Best Local Similarity 19.1%; Pred. No. 2e-06;

Matches 210; Conservative 146; Mismatches 381; Indels 362; Gaps 54;

Qy 25 VKGTVISSENEPLIGANVVVG-----NTTIGATDLDGNFTLSVPANAKMLRVSYSGM 79

Db 136 LKG-ILDSNGQVFLNPGITGKDAINTNGFTASTLD-----ISNENIKARPTFE 187

Qy 80 TTKEVAITANVMK-----IVLDPD-----SKVLEQVVVLGYGTGQKLSVSGSVAKVSSEK 129

Db 188 QTKOKALAEIVNHLITVKGKGSVNLIGGKVKNEGVI-----SVNGGSISLLAGQK 238

Qy 130 LAEPVAN-----IMDALQSQVAGMQVWTTSGDPTAVASVEIHGTGSLGASSAPLYVDG 184

Db 239 ITISDIINPTTYSIAAPENAEVNLGDIKPAKGGNINVRATIRNQGKLSAD----- 289

Qy 185 MQTSLDVAWNPNDPESMSVLKDSATSIYGAAA-----NGVVFIQTKGKMSERGIT 240

Db 290 -----SVSKDKSGNIVLSAKEGEAEIGGVISAQNOQAK---GCKL- 326

Qy 241 FNASYGISQILANTKPLDNMTGDE-----LLDFQVKAG---FWGNNQTVQVKOMILA 290

Db 327 -----MITGDKVLTGAVIDLJSGKEGGETYLGGDERGEGKNGIQLA 368

Qy 291 GAEDLYGNYD---SLKDEYKTLFPVDFNHDADWLKALFKTAPTSQDTSFSGG---SQG 344

Db 369 KKTSLKSGSTINVSKEKGRAIV-----W-----GDALIDGNINAQG 407

Qy 345 TSYASIGYFDQEGMAREPANFKRYSGRLNF--ESRI---NEWL-----KVGANLSGAIA 395

Db 408 SGDIATKG-----GFVETSGHDLFIKDNAIVDAKEWLLDFDNVSIANAETAGRS 455

Qy 396 RRSADYFGKYMGSGTFGLVTPRYNPFVNGDLADVYMYCATPSTPEPFACKRPF 455

Db 456 NTSDE---DEYTGSGNSASTPKRNEKKTTLTNTLESI---LKKGTFTVNIT---ANORIY 506

Qy 456 SSSHQANVNGFAQITPIKGLTL---KAAQAGVDITNTRTSSKRMPPNPYDSTPLGERRE 511

Db 507 VNSSINLS-NG-----SLTLWSEGRSGGVINNDITT-----GDDTR 543

Qy 512 RA-----YRDVSKSFTNTAYPEKSIDEKHDLTALMG--HEYIEYEGDVIGASSKGFES 562

Db 544 GANLTIYSGGWDVHKNISLGAQGNINITAKQDIAFEKGSNQVITCGQTITSGNQKGRF 603

Qy 563 DKMLLSQKGTGNSLSLPEHRVAEYAYLSFFSRFNYGFKWKMYIDFSVRNDQSSRFGSNN 622

Db 604 NNVSLL---NGTSGGLQFTTKRTNKYAITNKE-----GTLN 636

Qy 623 RSAWFYSVVGMPDI-----YNKFIQESNW-LSDLRLKMSYGTGTGSEIGNYNHQA 671

Db 637 -----ISGKNISMWLPKNESGYDKPKRITWNLTSLNV-----SESSEFN--- 677

Qy 672 LVTVNNYTEDAMGLSISTAGNPDLSWEKOSQFNFLAAGAFNNRLSAEVDFFVYRTTNDML 731

Db 678 -LTIDSRGSDSAGTLTQPYNLNGISFNKDTFNV-----ERNARVNF----- 718

Qy 732 IDVPMP-----YISGFFSQYQ-NVSGMKNTGVDLSL-----KGTIYQNKDWNVYAS 776

Db 719 -DIKAPIGINKYSSSLNYASFNGNISVSGGSDVFTLLASSSNVQTFGVWINSKYFNVSTG 777

Qy 777 ANFNANRQETIKLFFGLNKYMLPN-TGTIWEIGYPNSFYMAEYAGDKTKGKOLWYVPGQ 835

Db 778 SSLRFTKSGSTKTFPSIEKDLTUNATG-----GNITLLQVEGTDGMIKGI-VAKKN 828

Qy 836 VDADGNKVTTTQVYSADLETRIDKSVT-----PPTGGSFSLGASWKGL 877

Db 829 ITPFEGGNITFGSRKA--VTEIEGNVTINNANVTLLIGSDPDNHQKPLTIKKDVIINSGL 886

Qy 878 SLDADPAYIVGKMINND-----RYPTENAGGLMQLNKDKMLLNATWEDNKETDVPKLG 931

Db 887 TAGGNIVNIAGNLTVESNANFKAITNFTFNVGGLFD-----NKGNSNISIAKGG 935

Qy 932 QSPQFDTHLENASFLRLKNLKLTYVLPNSLFAQNVIGGARVYLMARL-----LTVT 985

Db 936 -----ARFKDIDNSKNLSITNTSSSTVYRTIISG-----NITKNGDLNIT 975

Qy 986 KYKGFDP--AGGNVGNQ 1002

Db 976 N-EGSDTEMQIGGDVSQKE 993

RESULT 10

US-08-719-641-2

; Sequence 2, Application US/08719641

; Patent No. 6218141

; GENERAL INFORMATION:

; APPLICANT: Barenkamp, Stephen J

; TITLE OF INVENTION: High Molecular Weight Surface Proteins

; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus

; NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/719,641

FILING DATE:

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/302,832

FILING DATE: 16-MAR-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US PCT/US93/02166

FILING DATE: 16-MAR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9205704.1

FILING DATE: 16-MAR-1992

ATTORNEY/AGENT INFORMATION:

NAME: Berketreiser, Jerry W

REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER: 1038-625

TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1536 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE:

US-08-719-641-2

Query Match 3.6%; Score 191; DB 3; Length 1536;

Best Local Similarity 19.1%; Pred. No. 2e-06;

Matches 210; Conservative 146; Mismatches 381; Indels 362; Gaps 54;

QY 25 VGTWISSDNEPLIGANVVVG-----NTTGAATDLDGNFTLSVPANAKMLRVSYSGM 79
DB 136 LKG-ILDSNGQVFLINPNIGITIGHDAIINTNGFTASTLD-----ISNENIKARNFTFE 187
QY 80 TTKVAIANVMK---IVLDPD-----SKVLEQVVVLGYGTGKLSVSGSVAKVSSEK 129
DB 188 QTKDALAEIVNHLITVGKQSVNLIGKVKNEGVI-----SVNGGSISLAGOK 238
QY 130 LAEPKPVAN-----IMDALQGVAGQVMTSGDPTAVASVEIHGTGSLGASSAPLYIVDG 184
DB 239 ITISDIINPTITYSIAAPENEAVALNIDIFAKGGINVRAATIRNQKLSAD-----289
QY 185 MOTSLDVVATMPNDFESMVLKDSATSIYGARAA-----NGVFIQTKGQMSRGRIT 240
DB 290 -----SVSKDKSGNIVLSAKEGEABIGVISAQNOQAK---GGKL- 326
QY 241 FNASVGISQILNTKPLDNMTGDE-----LLDFQVKAG---FWNNQTVQKVKDMILA 290
DB 327 -----MITGDKVTLTKGAVIDLSGKEGGTYLGGDERGEGKNGIQLA 368
QY 291 GAEDLYGNYD---SLKDEYKTLFFPVNFDHADLWALKFKAFTAPTSGDISFSGG---SQG 344
DB 369 KKTSLKSGSTINVSKEGGRIV-----W-----GDIALIDGINAQQ 407
QY 345 TSYIASIGVFDQEGMARPEANPKRYSGRINF---ESRI---NEWL-----KVGANLSGAJAN 395
DB 408 SGDIARTG-----GFVETSGHDLFIKDNAIVDAKEWLLDFDNVSNABTAGRS 455


```

QY 152 MTTSGDPTAVASVEIHGTGSLGASSAPLYIVDMQTSLDVATMNPNDP-----E 201
Db 79 LTNEGDNK--KGVSRGLD-----SSYTLILVDGKRVN-SRNAVFRHNDPDLNWIIPVDST 131
QY 202 SSVLKDASATSIYGARAANGVVFOTKKGKMSERGITFNASYGISOILNTKPLDNMT 261
Db 132 RIEVVR-GPMSSLYGSDALGGVNNITTKIGOKWGTVDVT----- 172
QY 262 GDELLDFQVKGAGFWGNQNTQVKQKMDILAGADLYGNYDSLKDEYKTLFPVDFNHDADW 321
Db 173 -----TIQHRD-----RGDTYNG----- 186
QY 322 LKALFKTAPTSQDTSFSGSGSGTSYIASIGYFD-----QEGWARPANPKRYSGRLNFS 377
Db 187 --QFTSGPLIDGVL-----GKAYGSLAKREKDDPQNSTTTDTGETPRIEG---FSS 234
QY 378 RINELWKVGNLSGAIANRRSADYFGKYMGSGTGVLTMPRYNPPFDVNGDLADVYNY 437
Db 235 R-----DGNVEFAWTPNQNHDF-----TAG----- 255
QY 438 GATRPSTMEPEYFAKMRPFSSSHQANVNGFAQITPIKGLTLKAQAGVDITNTRTSSKRP 497
Db 256 GFDRDRDSDSLDKNR-LERQNYVSHNGRDY-----GTSELKYGKVEN----- 301
QY 498 NNPYDSTPLGERERRRAYRVDVSKFTNTAYKFSIDKHDLTALMGHEYIEYEGDVIGASS 557
Db 302 KNFGNSSPI-----TSESTVDGKVTL-----PLTAI--NQFLTVCGE----- 337
QY 558 KGFESDKL--MLLSQKGTGNSLSLEPHRVAEYAYLSFFSRFNYGFDK-----MYIDPSV 610
Db 338 --WRHDKLSDAVNLGCGTSKTS-----ASQVAL-----FVEDEWRIFPEPLATTGV 382
QY 611 RNDQSRFGSNRNSAWFVSGVMFDIYNKFIQESNWLSDLR-----LKMSYGTGNSIGN 666
Db 383 RMDHETGEH-----WSPRAYLVYATDVTYVKGWATAPKAPSLQLSPDWTSNCRG- 437
QY 667 YNHQALVTNNYTEDAMGLSISTAGNPDLSEKQSQNFGL-----AAGAF 712
Db 438 -----ACKIVGSPDLKPETSSEWELGYMGEGWLEGVSSVTVP 478
QY 713 NNRLSAEVDVYRTNDMLIDVPMYISGP-----FSQYQNVGSMKQNTGVDLSL 761
Db 479 RNDVKDRIS--ISRTSDVNAAPGYQNFVGFETGANGRRIPVFS--YNNVNAKRNQGVETEL 535
QY 762 KGTIQNKDWNVYAGANFNYN--ROEITKLPFGLNKMPLNTGTIWEIGYPN-SFYMA-EY 818
Db 536 K--IPFNDKWL--SINTYNDGRVSN--GENK--PLSPLPHLALEDWFSFVSGHY 585
QY 819 AG---IDKKTGK-----QLWYVPGQVDADGNKVTTSQYSADLETRIDKSVTPPTITGFS 870
Db 586 TGQKRADSATAKTPGTYTIW-----NTGAAWQVTKDKLR-----AGVLNL 626
QY 871 G-----ASWK-GLSLDADPFIYVGMKMINND--RYF 898
Db 627 GDKTANGTLDMKPDLSRD--DYSY-----NEDGRRYF 656

```

RESULT 13
 PCT-US95-06994-9
 ; Sequence 9, Application PC/TUS9506994
 ; GENERAL INFORMATION:
 ; APPLICANT: Children's Hospital & Medical Center
 ; APPLICANT: University of Washington
 ; APPLICANT: Washington State University Research Foundation
 ; APPLICANT: TARR, PHILLIP I
 ; APPLICANT: BILGE, SIMA S
 ; APPLICANT: BESSER, THOMAS E
 ; APPLICANT: VARY JR, JAMES C
 ; TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
 ; STREET: SUITE 2800, 1420 FIFTH AVENUE

```

; CITY: SEATTLE
; STATE: WASHINGTON
; COUNTRY: USA
; ZIP: WA 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: 07-JUN-95
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/365,714
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BRODERICK, THOMAS F
; REGISTRATION NUMBER: 31,332
; REFERENCE/DOCKET NUMBER: CHOR-18591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 682 8100
; TELEFAX: (206) 224 0779
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 703 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: E. coli Cira protein amino acid sequence,
; DESCRIPTION: wherein "xaa" residues represent gaps
; DESCRIPTION: introduced to facilitate best alignment with
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: ESCHERICHIA COLI
; PCT-US95-06994-9

```

```

Query Match      3.4%; Score 178; DB 5; Length 703;
Best Local Similarity 19.5%; Pred. No. 6.4e-06;
Matches 165; Conservative 101; Mismatches 291; Indels 288; Gaps 43;

QY 92 IVLDPDSKVLQVVLGYGTGQKLTSTGSGVAKVSSEKLAEPVANIMDALQGVAGMQV 151
Db 24 LAVDDG--ETWVTVASSVEQNLKADAPASIVITQEDLQRPQNLKXDLK-EVPGVQ- 78
QY 152 MTTSGDPTAVASVEIHGTGSLGASSAPLYIVDMQTSLDVATMNPNDP----- 200
Db 79 LTNEGDNKGVKXKSIRGLD-----SSYTLILVDGKRVN-SRNAVFRHNDPDLNWIIXXPVD 133
QY 201 --ESMSVLKDSATSIYGARAANGVVFOTKKGKMSERGITFNASYGISOILNTKPLDN 258
Db 134 SIERIEVVR-GPMSSLYGSDALGGVNNITTKIGOKWGTVDVT----- 177
QY 259 NMTGDELLDFQVKGAGFWGNQNTQVKQKMDILAGADLYGNYDSLKDEYKTLFPVDFNHD 318
Db 178 -----TIQHRD-----RGDTYNG----- 191
QY 319 ADMLKALFKTAPTSQDTSFSGSGSGTSYIASIGYFD-----QEGWARPANPKRYSGRLN 374
Db 192 -----QFTSGPLIDGVL-----GKAYGSLAKREKDDPQNSTTTDTGETPRIEG--- 236
QY 375 FESRINELWKVGNLSGAIANRRSADYFGKYMGSGTGVLTMPRYNPPFDVNGDLADVY 434
Db 237 FSSR-----DGNVEFAWTPNQNHDFTAGY-----GFDRQDRDSDSLDKNXXXXXXX 282
QY 435 YMYGATRPSTMEPEYFAKMRPFSSSHQANVNGFAQITPIKGLTLKAQAGVDITNTRTSSK 494
Db 283 XXXRLERQNY-----SVSHNGRDYGTSELKYGKVENKXXXXXXX 330
QY 495 RMPNNPYDSTPLGERERRRAYRVDVSKFTNTAYKFSIDKHDLTALMGHEYIEYEGDVIG 554
Db 331 XXPGN---SSPI-----TSESTVDGKVTL-----PLTAI--NQFLTVCGE--- 366

```

QY 555 ASSKGPESDKL--MLLSQKGTGNSLSLPEHRVAEYAYLSFFSRFNYGFDKW-----MYID 607
Db 367 -----WRHDKLSDAVNLGTGSSKTS-----ASQYAL-----FVEDEWRIFEEPLALT 408
QY 608 FSVNRDQSSRFGNNRSANFYSGVGMPIYKFKTQESNWLSDLR-----LKQSYGTTGNSE 663
Db 409 TGVMDDBHETGEGH-----WSPRAYLVYNATDVTVKGMWATAFKAPSLQLQSPDWTNSNC 464
QY 664 IGNYNHQALVTNNYTEDAGLSISTAGNPDLSWEKQSQFNGL-----AA 709
Db 465 RG-----ACKIVGSPDLKPETSESWELGLYMGEGWLEGVESV 504
QY 710 GAFNNRLSAEYDFVVRTNMDLIDVPMPIISGR-----FSGYQNVGSMKNTGVD 758
Db 505 TVFRNDVKDRIS-----ISTSDVNAAPGYQNFVGFETGANGRRIPVFS--YNNVKARKOGVE 561
QY 759 LSLKGTIYQKNDVNASANFNYN--ROEITKLFGLNKM--LPNTGTIWEIGYPSNPFYM 815
Db 562 TELK--IPFNDEWKL--SINVYNDGRDVS--GENKPLSLDP-----FHL 601
QY 816 AEVAGIDKTKGQLW--YVPGQVDADGNKVTTSQYSADLETRDKSVTPPITGGPSL--- 870
Db 602 A-----LEDWSPFVSGHY-----TGQKRAD-----SATAKTPGTYTIWNT 636
QY 871 GASWK 875
Db 637 GAUWQ 641

RESULT 14

US-09-328-352-7650
; Sequence 7650, Application US/09328352
; Patent No. 8562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7650
; LENGTH: 759
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7650

Query Match 3.3%; Score 175.5; DB 4; Length 759;
Best Local Similarity 18.7%; Pred. No. 1.2e-05;
Matches 170; Conservative 118; Mismatches 304; Indels 317; Gaps 41;
QY 101 LEQVVVLGYGTGOKLSTVSGSVAKVSSEKL-----ABKPVANIMDALQGVAGQVNMVTT 155
Db 52 LQKIVVATATPKNIAEAGTQVSIQKQIQQATAGRKVADILAQVLPSLASSGTTSN 111
QY 156 GDPTAVASVEIHGTSGASAPLYVDGM--QT--SLDVATMN-----PNDFESMVLKXAS 210
Db 112 -----YQGTMRGRN--VLVMIDGVSQTGRDVSRLQNSISPGMIERIEVI--SG 156
QY 211 ATSIYGARANGVVFIOTKGKMSERGITFNASYGISQILNTKPLDNMMTGDLELLDFQV 270
Db 157 ATSIYGATGGINIITKADTS-----KP-----LSFET 187
QY 271 KAGFWGNQTVQVKDMLAGAEDLYGNYSKDEYKTLFPVDFNHDADWLKALFKTAP 330
Db 188 KVGI-----TSSDTFRS--DGLAVEVGQS--VSFN-----213
QY 331 TSQGDISFGSGSQTSYASIGYDFQEG--WAREP-----ANFKYSGLNPFESRINE 381
Db 214 --KGNMD--GFLGANFTSRGSPQDNGDRISLSPWQSGTMDTDTIDVNGRLNLFNDTQ 268
QY 382 WLKYGANLSGAIANRRSADYFGKY--YM-----GSGTFCVLTMPRY-- 420

Db 269 TLSFGAQY---YKDKQDTDGPDYSYLPSTTSKSNDAATPTTYKAIGLKLNSPLFTERYAV 325
QY 421 ---YNPFDVNGDLADV-----YYMVGATRPSMTPEYFAKMPFSSSESHQANVNGF 467
Db 326 NSQYQNDQFLGQILNVEAYYRNEKSRFFPYGLSNKSVT-----SVNQ 368
QY 468 AQITPTKGLTLKRAQAGVDITNTRTSSKRMPPNPD-----STPLGE 508
Db 369 QSEIEVAGLRSTRWOTDLNAN--RDMKITTYGLDYDWEKDKQFVDILATQYPLYVYPTGQ 426
QY 509 RRERAYRVDYSKTPNTAAYKFSIDEKHEDETALMGHEYIEYEGDVI GASSKGFSDKMLML 568
Db 427 RKGYGENTEIQNIGAFVQSDYAVTDKLNQAGIRYQYIQADTDAY-----IPSRETTMV 480
QY 569 SQKGTGNSLSLPEHRVAEYAYLSFFSRFNYGFDKWYIDFSVRNDQSSRFGSNRSANFY 628
Db 481 PAGSTHDDKPL-----FNLGA-----VY 498
QY 629 SVGGMFDIYNKFTQESNWLSDLRLKMSYGTGTGNSIGNYHQALVTNNYTEDAGLSIS 688
Db 499 KLTDAAQVYANFSQGSFDPVQRM-----LRDVSITYTUSTANLOPI 539
QY 689 TAGNPDLW--EKQSQNFGLA-----AGAFNNRLSAEVDYFYRTTN--DMLIDVP-- 735
Db 540 TVNSYELGWLNRQDDGLNLGLTGFTYNTSKTVQFNRAAKVVDTDQKVYGAETISYPPM 599
QY 736 -----MPYISGFPFSQYQNVGSMKNTGVDLSLKGITIQNKDWNVYASANFNYNQEI 787
Db 600 ENYKVGGLTYTRGQYKDVANKWHELNSFTVAPVKGLTF--AEWD-----NNEGYGVRVQM 653
QY 788 KLFFGLNKYMLPNTGTIWEIGYPSNFYMAEYAGIDKTKGQLWYVPGQVDADGNKVTTSQ 847
Db 654 QAIKGNK-----AYKDDRELAAPATTQDEAFRK-----AVENDANSA--- 691
QY 848 YSADLETRIDKSVTPPITGFGSISGASWKGSLDLPADFAIYVGMKMINNDRYFT----- 899
Db 692 -----AQIKGYTMTDVLAHF---PAWKG---RVDF---GVYNVWNRQYRTVPAQQA 735
QY 900 ENAGGLMQL 908
Db 736 SNANPLLA 744

RESULT 15

US-09-252-991A-30809
; Sequence 30809, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30809
; LENGTH: 756
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30809

Query Match 3.3%; Score 174; DB 4; Length 756;
Best Local Similarity 22.3%; Pred. No. 1.5e-05;
Matches 176; Conservative 97; Mismatches 302; Indels 214; Gaps 44;
QY 59 GNTFL--SVPANAKMLRVSYSGMTTKEVAJANYMKIVLDPDSKVLQVVLGYGTGOKLS 116
Db 11 GNLMPRSIPLRPAPLALSLS-----LFASFSAPALAADP---VEQQMVVIGSRAPTRIS 62

Search completed: January 7, 2004, 19:04:17
Job time : 29 secs

THIS PAGE BLANK (USPTO)

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	201	3.8	2057	12	US-10-417-280A-2		Sequence 2, Appli
2	197	3.7	1536	12	US-10-193-764-63		Sequence 63, Appli
3	192.5	3.6	682	12	US-10-238-075-1077		Sequence 1077, Ap
4	191	3.6	1536	14	US-10-092-880-2		Sequence 2, Appli
5	173.5	3.3	1600	14	US-10-092-880-10		Sequence 10, Appli
6	171	3.2	2122	10	US-09-813-214B-9		Sequence 9, Appli
7	168	3.2	767	12	US-09-882-227-78		Sequence 78, Appli
8	168	3.2	1599	14	US-10-092-880-9		Sequence 9, Appli
9	168	3.2	1849	12	US-10-369-493-18460		Sequence 18460, A
10	167.5	3.2	2834	12	US-10-085-959-252		Sequence 252, App
11	167	3.2	1974	10	US-09-895-913A-12		Sequence 12, Appl
12	164.5	3.1	2732	12	US-10-238-075-1119		Sequence 1119, Ap
13	162.5	3.1	1371	12	US-10-238-075-891		Sequence 891, Appl
14	160	3.0	2893	12	US-09-882-227-522		Sequence 522, App
15	159.5	3.0	3241	9	US-09-841-786-1		Sequence 1, Appli

```

143 G-----LTTNLSDPITGCHYENHNGYFVYIDASCKQVTGLQNDGNLQYFD----- 189
197 PNPESMVLKDSATSIIY-----CARAANGVVFQTK-----XGKMSERGRITFNASYG 246
190 DNGYQVAGSPRDVNGKRIYFDSVTGKASSNDVINGRAQGYDAQGNQLKESYVADSGQT 249
247 ISQILNTKPLDNMTGDELDF-----QVKAGPWG--NNOIVQVKDM--ILAGAEDLYG 297
250 YFPDNGCPILIGITIDGNLOYNQOQGVJGKGFQDVNNKRIYFAPNTGNAVANTEILNG 309
298 -----NYDSLKDEYKTLFPVDFNHDWLKALFKTAPTQSGDISFSGSGSTSYAS 350
310 KLQGRDANGQVKNAPSKDVAGNTFYFDANGVML-----TGLQTIS-----GKTY--- 354
351 IGYPDQGMAREPANFKRYSGRLNPFESRINEWLKVGANLSG-----AIANR 396
355 --YLDQGHLR-----KNYAGTFN-----NQMYFDADTGAGKTAIEYQPDQGLVQSNE 402
397 RSADYFCKYMGs---GTFGLTWPVRYNPDV--NGDLADVYMYGATRPSTPEYPAK 451
403 NTHNAAKSYDKSFENVGVLTDADTWYRPTDILKNGD-----TWASTE---TD 449
452 MRP-----FSSSHQANVNGFAQITPIKGLTLKAQAGVDITNTRSSKRMNN----- 499
450 MRPLMTWPDKQTOANYLFMS-----SKGL-----GITTYTAATSOKTLNDAAFVIQT 500
500 PYDSTPIGERERRAYRVSFTWTAB--YKFSIDEKHDTALMGHYIYEGDVGIAGS 557
501 AIEQOISLKXSTELRDAIDSFKVTQANWNKQTEDEAFDGLQWLOQGGFLAYQDD-----SH 556
558 KGPESDKMLLSQKGTGNSLSLPEHRAVEAYLISFFRPNYGPDKWYIDFVSVRNDOSSR 617
557 RTPNTD-----SGNNRKLGRQPI-----NIDGS---KUTDGG 585
618 FGSNRNSAWFYSVCGMEDIYKFKIQ--ESNWLSDLRLKMSYGT--TGNSEICNTYHQAULT 674
586 KGSE-----FLANDIDNSPIVQAEQLNLHVL---MNFSGITGNNDNANFDGIRVDA 636
675 VNNYTEDAMGLS-----ISTAGNPDLS--WE 698
637 VDNVDALLKIAGDYPRKALYGTOKSDANANKHLSILEDWNGKQDPQVYVQNGNAQLTMDYT 696
699 KQSQFNGLAAGAFNRLS-----AEVDYFVR--TNDMLIDVPYPIISGFPSQYQNTGSMK 753
697 VTSQFGNSLTHGA--NRSRNMWYFLDTGYLLNGDLNKKIVDKNRPNSGTLNRIANGSDTK 755
754 -----NTGVDLSLKGTIYQNKDMNYYASANFNRYNRQEIITKL 789
756 VIPNYSFVRAHDYDAQDPIRKAMIDHGIKNMQDPTTFDQ---LAQGMFYYKQDNPSG 812
790 FFLGKNTMLNPTGITIWEIGVNSPYMAEYAGI--DKKTGKOLWYVPQVDADGNK----- 842
813 FKXNDYNLPSA-----YAMLTNKTDPVPRVY--GDMYLEGGQYMEKG 854
843 -VITSQVSADLETFRIDKSVTPITGFSGLASWKLGLS---DADFAYIV--GKWMINNDR 896
855 TIYNPVSALLKARIKY-----VSGQGTMAFDSGKDLKGETDILLTSVRFPGKIMTSDQ 909
897 YFTE-----NAG-GL-----MQLNKDKML-----LNAW 918
910 TTTQDSQDYKNQGIYVGNPNLKLNDKTIITLHGKAHKNQLYPALVLSNDSGIDVY 969
919 TENKETVDVKLGOSPOFDTHLENASFLRLKMLKLYVLPNSL-----FAGQNVIGGA 972
970 DSDDK---APTARTNDNGDLIFHKTNTFVKQDGTIINYEMKGSINALISGLYGVVYVPGA 1026
973 RVLVARNLLTVTK-----YKGF-----DPEAGNV 998
1027 SDSQARTVATESSSNDGVSFTHSNAALDSNVIEYGFNSFOAMPTSPSEQSTNV 1079

```

```

US-10-193-764-63
; Sequence 63, Application US/10193764
; Publication No. US20030133943A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
; FILE REFERENCE: 1038-1239MIS
; CURRENT APPLICATION NUMBER: US/10/193,764
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/167,568
; PRIOR FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 63
; LENGTH: 1536
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-193-764-63

Query Match 3.7%; Score 197; DB 12; Length 1536;
Best Local Similarity 18.7%; Pred. No. 1e-06;
Matches 206; Conservative 151; Mismatches 377; Indels 368; Gaps 51;

QY 25 VKGTVISSEDEPLICANVVVVG-----NTTGAATDLDGNFTLSVPANAKMLRVSYSGM 79
DB 136 LKG-ILDSNGQVFLINPNGITIGKAIINTNGFTASTLD-----ISNETIKARNPTFE 187
QY 80 TTKEVAIAVMK-----IVLDPD-----SKVLEQVVVLGVGTGQKLSTVSGSVAKVSSEK 129
DB 188 QTKDALAIIVHGLITVGKDSVNLIGKVKNEGI-----SVNGSISLLAGQK 238
QY 130 LAEKPVAN-----IMDALQQVAGQVMTTSGDPTAVASVEIHGTGSLGASSAPLYVDG 184
DB 239 ITISDIINPTITVTSIAAPEANVLGDIPAKGNINVRAATIRNQKLSAD----- 289
QY 185 MQTSLDVVATMPNDFESMVLKDSATSIIYGARAA-----NGVVFQTKKGMSEGRIT 240
DB 290 -----SVSKDSKGNIVLSAKEGEARIGGVISAQNOQAK---GGKL- 326
QY 241 FNASYGISQILNTKPLDNMTGDE-----LLDFQVQAG---FMGNNTQVQVKDMILA 290
DB 327 -----MITGDKVTLTKGAVIDLKSGEGGETYLGDERGEGKGIQLA 368
QY 291 GAEDLYGNYD---SLKDEYKTLF-----PVPFNHDADWLKALFKTAPTQSGDISFSGS 342
DB 369 KKTSLKSGSTINVSGKEGGRIVMGDIALIDGNINAQ-----GSGDIKTKGGF 417
QY 343 QGTS-----YYASIGYPDQGMAREPANFK---RYSGRINPFESRINEWLKVGANLSGAIA 395
DB 418 VETSGHDLFIKONAIYDAKEWLLDPDNVSINAEATAGRSN--TSEDDYTGSGNSASTPKRN 476
QY 396 RRSADYFG--KYVMGSGTFFVLTPRYNPNYFPVNGDLADVYMYGATRPSTPEYPAKM 452
DB 477 KEKTLTNTLSILKKGIFVNI-----ANQ 503
QY 453 RPFSSSHQANVNGFAQITPIKGLTL---KQAGVDITNTRTSSKRMNNPYDSTPLGE 508
DB 504 RIYVNSSINLS-NG-----SLTLWSEGRSGGVIEINNDITT-----GD 540
QY 509 RRERA-----YRDVSKSPTNTABYKFSIDEKHLDTALMG-HEVIEYEGDVIGASSKG 559
DB 541 DTRGANLTIYSGWVDVHKNISLGAQGNINITAKQDIAFERKGSNQVITQGTITSGNQKG 600
QY 560 FESDKMLLSQKGTGNSLSLPEHRAVEAYLISFFRPNYGFDKWYIDFVSVRNDOSSRFG 619
DB 601 FRFNVL---NGTSGLOFTTKRTNKYAITNKE-----G 633
QY 620 SNRNSAWFYSVCGMFDI-----YNKFTQESNW--LSDRLKMSYGTGTGNSEIGTNYN 668
DB 634 TLN-----ISGKVNISMVLPKNESGYDKFKRTYWNLTSLNV-----SESSEFN 677

```



```

136 LKG-ILDSNGQVFLINPNTIGKDAIINTNGFTASTLD-----ISNENIKARNFTPE 187
QY 80 TKEVAIANVWK-----IVLDDP-----SKLEQVVVLGYGTGQKLSVSGSVAKUSSEK 129
Db 188 QTKDKALAEIVNHLITVGDGKSVNLGGKYNKEGVI-----SVNGGSIALLAGQK 238
QY 130 LAEPKPVAN-----IMDALQGVAGQVMTTSGDPTAVASVEIHGTGSLGASSAPLYIVDG 184
Db 239 ITSDIINPTIYSIAAPEANVNLGDIPAKGGNINVRATIRNQKLSAD-----289
QY 185 MOTSLDVVATMNPDPFESMYSVLKASATSIYGARAA-----NGVVFIQTKGKMSERGRIT 240
Db 290 -----SVSKDKSGNIVLSAKEAGEIGVISAQNOQAK---GGKL- 326
QY 241 FNASYGISQILNTPKLDNMGTGB-----LLDFQVQAG---FWGNQNTVQVKOMILA 290
Db 327 -----MITGDKVTLTKTGAVIDLSGKEGETYLGDERGEGKNGIOLA 368
QY 291 GAEDLYGNYD---SLKDEYKTLFPVDFNHDADWLKALPKTAPTSGQDISFSGG---SQG 344
Db 369 KKTSLKSGSTINVSCKEKGRAIV-----W-----GDIALIDGNINAOQ 407
QY 345 TSYVASICYFDQEGWAREPANFRKYSGRNLF--ESRI---NEWL-----KVGANLSGAIAN 395
Db 408 SGDIAGTG-----GFVETSGHDLFIKDNAIVDAKEWLLDPDNVSNIAETAGRS 455
QY 396 RRSADYFGKYMGGTGVLTMPRYNPPDFVNGDADVYMYGATRPMSMTEPYFAKWRPF 455
Db 456 NTSED---DEYTGSGNSASTPKRNEKTKLTNTTLESII--LKKGTFFVNIT---ANQRIY 506
QY 456 SSESQANVNGFAQITPIKGLTL---KAQAGVDITNTRTSSKEMPNPYDSTPLGERRE 511
Db 507 VNSSINLS-NG-----SLTMSGRSGGVGEINNDIT---GDTR 543
QY 512 RA-----YRDVSKSFTNTAETKFSIDEKHDLTALMG--HEYIEYEGDVIGASSKGFE 562
Db 544 GANLTIYSGGVVDVHKNSILGAQGNINITAKQDIAPEKSGNOVITGOTTISGKQGRF 603
QY 563 DKMLLSGGTGNLSLPEHVRVAYLSPFRNRYGDKMYIDFVRNDQSRFGSN 622
Db 604 NVSL---NGTSGSLQPTTKRTNKYAITNKPE-----GTLN 636
QY 623 RSAMFYSGGFMFI-----YNKEIQESNW--LSDLRKMSYGTGTGSEIGNYHQA 671
Db 637 -----ISGKVNISMVLPKRESYDVKPKGTYWNLTSLNV-----SESGEFN--- 677
QY 672 LVTNNYTEDAMGLSISTAGNPDLSEKQSFNGLAAGAPNNRLSAEVDVPYVRTNDML 731
Db 678 -LTIDSRGSDSAGTLTQPNYNLNGISFNKDTFFNV-----ERNARVNP-----718
QY 732 IDVPMP-----YISGPFSSQYO--NVGSMKNTGVDLSL-----KGTIYONKDMNYYAS 776
Db 719 -DIRAPIGINKYSLNVAENFNGNISVSGGSDVDPTLLASSNVQTPGVWINSKYFNVSTG 777
QY 777 ANFNYNRQELTKFGLNLYMLPN--TGTWEIGYPNFPYMAEYAGIDKTKQKLYWYPGQ 835
Db 778 SSLRFTKSGSTKTFSEKIDJTLNATG-----GNITLLQVEGTGMIGKGI--VAKKN 828
QY 836 VDADGNKVTTSQYSGADLETRIDKSVT-----PPITGGFSLGASWKL 877
Db 829 ITPEGGNITPGRKA--VTEIEGNTVINNNANVTLGSDPDHQQPLTIKDVILNSGNL 886
QY 878 SLOADPFIYVGVKMINND-----RYFTENAGGLMQLNKKDKMLLNAMTENDKTDVPLG 931
Db 887 TAGGNIVNIAGLAVESNANFAITNFTFNVGGLFD-----NKGNSNISIAKGG 935
QY 932 QSPQDTHLENAFLRKNLKLTVLPNSLFAQONVIGGARVYLMARNL-----LTVT 985
Db 936 -----ARPKOIDSNKLSITNTSSSTYRTIISG-----NITKNKGNLDNIT 975
QY 986 KYGFDPE--AGGNVGNQ 1002

```

```

Db 976 N-EGSDTEMQIGGDVSKQE 993

RESULT 5
US-10-092-880-10
; Sequence 10, Application US/10092880
; Publication No. US20020164354A1
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE
; FILE REFERENCE: HAEMOPHILUS
; CURRENT APPLICATION NUMBER: US/10/092,880
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: 09/155,614
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 08/617,697
; PRIOR FILING DATE: 1996-04-01
; PRIOR APPLICATION NUMBER: PCT/US97/04707
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-092-880-10

Query Match 3.3%; Score 173.5; DB 14; Length 1600;
Best Local Similarity 19.1%; Pred. No. 9.8e-05;
Matches 228; Conservative 161; Mismatches 448; Indels 355; Gaps 56;

QY 16 GWAMAQNRIV-KGTVISSDNEPLIGANVVVGVNTTIGATDLDGNTLSVPANAK---M 71
Db 364 GIQAKKTKLEKSGSTINVSKEK--GGRAIVWDGIAL-----IDGINAOGSDIAKTGGF 416
QY 72 LRVYSGMTTKVAIANVWKIVLDDPKSVKLEQVVLGYGTGQKLSVSGSVAKVS--SEK 129
Db 417 VETSGHDLISGDDVIIVDAKEWLLDPDDVSIETLTSGNNTGENQGYTTGDTKESPKGNS 476
QY 130 LAEPKPVANIMDALQGVAGQVMTTSGDPTAVASVEIHGTGSLGASAPLYI---VDGMQ 186
Db 477 ISKPTLTN--STLE-----QILRRGSYVNITANNRIYVNSSINLSNGSLTLHTKRDGVK 528
QY 187 TSLDVVATMNP-----DFESMVLKASATSIYGARAAVGVFIQ 227
Db 529 INGDIITSNENGNTIKAGSWVDVHKNTILGTGFLNI--VAGDSVAFEREKDKARNAT--- 583
QY 228 TKGKMSERGRITFN-----ASYGISQILNTPKLDNMNTGDELLDFQVKA 272
Db 584 --DAQITTAQTTITVKKDDKQFRFNNVSLNGTGKLFANQNNPHTKFDGEINI-----S 636
QY 273 GPMGNQNTVQK--VKDMILAGAEDLYGNYDSL-----KOEY-----306
Db 637 GIVTINQTKKDVK--YMNASKDSYWNVSLSLTNTVQKPTFIKFDVDSGNSQDLRSRRS 694
QY 307 -----GKTLFPVDFNHDADWLKALPKTAPTSGD-----ISPSGSGQTSYYAS 350
Db 695 FAGVHFNGIGGKTFNFIKAN-----AKALFKLPNAATDPKBLPTFFNANITAT----- 744
QY 351 IGYFDQEGMAREPANFRKYSGRNLFESRINEWLVKANLSGAIANRRSADYFGKYVMGSG 410
Db 745 -GNSDSVMPDIHANLTSRAAGINMDS-----INITGGL-----777
QY 411 TPGVLTMPRYNPPDFVNGDL---ADVYMYGATRPMSMTEPYFAKMRPFSS-----457
Db 778 DFSITSHNRNSNAFEIKKDLTINATGNSFSLKQTKDSFYNEY--SKHAINSSHNLTLILGNN 836
QY 458 -----ESHQANVNGFAQITPIKGLTLKAQAGVDITNTRTSSKEMP-----497
Db 837 VTILGGENSSSITNTNITNKANVTLOA-----DTSNNTGLKKRTLTTLGNISVEGNLSLT 892
QY 498 -----NPNYDSTPLGERRERAYRVDVSKSFTNTAETKFSI-----531

```



```

; Publication No. US20030158396A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20030158396A1 Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/047002
; CURRENT APPLICATION NUMBER: US/09/882,227
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/902,615
; PRIOR FILING DATE: 1997-07-29
; NUMBER OF SEQ ID NOS: 638
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78
; LENGTH: 767
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 304
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-882-227-78

Query Match      3.2%; Score 168; DB 12; Length 767;
Best Local Similarity 19.2%; Pred. No. 8.5e-05;
Matches 172; Conservative 118; Mismatches 341; Indels 264; Gaps 40;

Qy 101 LEQVVLGYGT-----GQLSTVSGSVAKVSSEKLAEPVANIMDALQGVAGMOV 151
Db 23 LERVEASGVANDKEAPLSWKSKEVRNVTGRTVSNKQLTKSANQSEALQ-NVPGVHI 81

Qy 152 MTTSGDPTAVA-SVEIHGTGSLGASSAPLYIVDG-----MOTSLDVATMNPDPES---M 203
Db 82 RNATGIGAVPSFSVRGFGSSGSHNTAMVLVNGIPIYVAPYVDISIPFPVTFQSDVRI 141

Qy 204 SVLKDAATSIIYARAANGVVFIOTK-----KGMSEGRITFNASYGISQILNTKPLD 257
Db 142 SVTKGGSVSR-YGPNVFGVINVTIGTPTKWSQVSRATFWGKSENGGFFNQNSKNLD 200

Qy 258 NMWTGDELLDPQVKGAGFWGNQTVQVKYKMDLAGAEDLYGNYDSLKDEYKTLFPVDFNH 317
Db 201 KSLANNMLFDLYLRTGGMNKHFGIOAANWLKQGGFRYNSPTNIQNYMLDSLQI---N 257

Qy 318 DADWLKALFKTAPTSGDIDISFGSGSGTYSYASIGYF--DOEGMAREPANFKRYSGRLL-- 373
Db 258 DSNKITAFFQ-----YY---NYFMADPGSLGIEAYNQNRFPQNNRPN 295

Qy 374 -NFESRINELWKGANLSGAIANRRSADYFGKYMGSGTGTGVLTPRYPYPPFVNGDLA- 431
Db 296 NNSKGRAKXGAVYQNFPG-----DTDKIG-----GDFTFS-----YYG-HDMSRDFQF 338

Qy 432 DVYMYCATPSMTEPY-----FAKMRPFSSSHOANVNGFAQITPIK-----GL 476
Db 339 DSNFLVNTNPKGFPVTDQNYPGFFIFDLHRRYIMNAFNLVNTNPKVQTFENVGM 398

Qy 477 TLXA-----QAGVDITNRTSSKRNPNNDSTPLGERRERAYRDVSKSFNTNABYK 528
Db 399 RFWTMDMYFRLDQSTCEKTDIFNGVCRMP--PF-----VLSKSPNNQNL- 441

Qy 529 PSIDEKHDLTALMGHEVIEVEGDVIGASSGKFESDKMLLSQGTGNSLSLPEHRVAEYA 588
Db 442 -----FNNYTAVWLSDKIEL-----FDSKLVIITPG-----LRYT 470

Qy 589 YLSPFSFRNYGDFDKWYIDFSVRNDQSSRRFGSNRRSAWFYSGGMPDIYNKFIQESNWL 648
Db 471 FL-----NYYNKEPEKHDFSVNMTTKR---QNEWSPALNIG-----YKPMENWIW 513

Qy 649 DLRLKMSYGTGTGNSIGNYHQALVTVNNTYEDAMGLSISTAGNPDLISWEKQSOFPGLA 708

```

```

Db 514 YANYRRSFIPPOHTMLG-----ITRTNYNQIENIEVQQR-----YSYKNLLSPN---- 558
Qy 709 AGAFNNRLSAEVDVYVRTTNDMLIDVPMPIYISGFFSQYQNVGSMKKTGVDSLKGTIYQN 768
Db 559 -----TNYFVIFAKRYIAGYSQPINA-----RSQGVELELYIYAPIRG 597
Qy 769 KQWNV---YASANFNYNROEITKLFGL-NK-----YMLPN-----TGTWE 806
Db 598 LQPHVAYTIIDARITSNADDAIYFVIGVKNKPDIDKGLPYVSPNQIFDMMYTKHTT 657
Qy 807 IGVPSFYMAEYAGIDKTKGKQLWYVPGQVADGNKVTTSQYSADLETRIDKSVTPPITG 866
Db 658 FGTSYFYSRAYSSMLNQAQKSTVCLP-----LNPEYTG 691
Qy 867 GFSLGASWKGSLSLDADPAY-----IVGKWMIN---NDRYFTENAG 903
Db 692 GLEYGCSVGL-LPLYFVLNVQSSVLMQSGRHKITGSLQINNLFNNKYFRGIG 745

RESULT 8
US-10-092-880-9
; Sequence 9, Application US/10092880
; Publication No. US20020164354A1
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE
; TITLE OF INVENTION: HAEMOPHILUS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/092,880
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: 09/155,614
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 08/617,697
; PRIOR FILING DATE: 1996-04-01
; PRIOR APPLICATION NUMBER: PCT/US97/04707
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 1599
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-092-880-9

Query Match      3.2%; Score 168; DB 14; Length 1599;
Best Local Similarity 19.8%; Pred. No. 0.00028;
Matches 231; Conservative 167; Mismatches 454; Indels 324; Gaps 59;

Qy 16 GWAMAQNRTV-KGTVISSEDNELPIGANVVVGNNTTIGAATDLDGNFTLSVSPANAK---M 71
Db 364 GIOLAKTKTLEKSTINVSGKEK--GGRALVNGDIAL-----IDGINAOGKDIATGGF 416

Qy 72 LRVSYSGMTTKEVAIANVMKVLDPDSKVLQVVLGYGTGQKLSVSGSVAKVSEK-- 129
Db 417 VETSGHYLSIDONAIKTKLEWLLDPENVITAPSASRVELGADRNHSAEVIKTLKKN 476

Qy 130 -----LAEPVANIMDAL-----QGOVA-GMOV--- 151
Db 477 TSUTTLNTTINLLKLSAHVNIITARKULTVNSSISIERGSHLILHSEGGQGVQIDKD 536

Qy 152 MTTSGDPTAVAS---VEIHGTGSLGASSAPLYIVDG-----MOTSLDVATM 195
Db 537 ITSEGGNLIYSGGWVDVHKNTITLGSGLNITTKEGDIAFEDKSGRNNLTITAGGITSG 596

Qy 196 NPNDPESMSV-----LKDASATSIYGARAANGVVVFIOTKKGKMS 234
Db 597 NSNGFRNNSVLSLGGKLSFTDSRDRGRRTKGNISNKFDPGLNISGTVDISMKAPKVS 656

Qy 235 -----EGR-----ITNASYGISQILNTKPLDNMTGDELLDPQVKGAGFWGNQTVQVK 285
Db 657 WFYRDKRGTWNVTTLNVTSGSKFNISIDSTGSGTSPSIRNAEL-----NGITFNKAT 710

Qy 286 DMLAGAEDLYGNYDSLKDEYKTLFPVDFNHDADWLKALFKTAPTSGDISFS- 344

```


Db	1509	SI-SLSPNSYDFCTGLDVSFKPLISWATSPASDVVVVTDTRTGSAKWPVSVQAQDLKG	1567
Qy	837	-----DADGNKVTTS---QYSADLETRIDKSVTPITGGFSGASWKGLS	878
Db	1568	LTNNNLASLYFFKDXGSKVITSDALQIYAN-----TSPTTGTGFKLNQNNWSTS	1617
RESULT 10			
US-10-085-959-252			
; Sequence 252, Application US/10085959			
; Publication No. US20030165870A1			
; GENERAL INFORMATION:			
; APPLICANT: Blattner, Frederick R.			
; APPLICANT: Welch, Rodney A.			
; APPLICANT: Burland, Valerie D.			
; TITLE OF INVENTION: No. US20030165870A1el Sequence of E. Coli CFT073			
; FILE REFERENCE: 960296.97648			
; CURRENT APPLICATION NUMBER: US/10/085.959			
; CURRENT FILING DATE: 2002-03-01			
; PRIOR APPLICATION NUMBER: 60/242,412			
; PRIOR FILING DATE: 2000-10-19			
; NUMBER OF SEQ ID NOS: 255			
; SOFTWARE: PatentIn version 3.1			
; SEQ ID NO 252			
; LENGTH: 2834			
; TYPE: PRF			
; ORGANISM: Escherichia coli			
; FEATURES:			
; NAME/KEY: misc feature			
; LOCATION: (493)..(493)			
; OTHER INFORMATION: The 'Xaa' at location 493 stands for Asn, Asp, His, or Tyr.			
; NAME/KEY: misc feature			
; LOCATION: (1158)..(1158)			
; OTHER INFORMATION: The 'Xaa' at location 1158 stands for Asp, or Tyr.			
; NAME/KEY: misc feature			
; LOCATION: (1159)..(1159)			
; OTHER INFORMATION: The 'Xaa' at location 1159 stands for Leu.			
; NAME/KEY: misc feature			
; LOCATION: (1168)..(1168)			
; OTHER INFORMATION: The 'Xaa' at location 1168 stands for Ser, or Asn.			
; NAME/KEY: misc feature			
; LOCATION: (1270)..(1270)			
; OTHER INFORMATION: The 'Xaa' at location 1270 stands for Thr, or Pro.			
; NAME/KEY: misc feature			
; LOCATION: (13480)..(13480)			
; OTHER INFORMATION: Unsure			
; NAME/KEY: misc feature			
; LOCATION: (31038)..(31038)			
; OTHER INFORMATION: Unsure			
; NAME/KEY: misc feature			
; LOCATION: (31042)..(31042)			
; OTHER INFORMATION: Unsure			
; NAME/KEY: misc feature			
; LOCATION: (31770)..(31770)			
; OTHER INFORMATION: Unsure			
; NAME/KEY: misc feature			
; LOCATION: (31799)..(31799)			
; OTHER INFORMATION: Unsure			
; NAME/KEY: misc feature			
; LOCATION: (44922)..(44922)			
; OTHER INFORMATION: Unsure			
US-10-085-959-252			
Query Match 3.2%; Score 167.5; DB 12; Length 2834;			
Best Local Similarity 18.8%; Pred. No. 0.00079;			
Matches 228; Conservative 147; Mismatches 384; Indels 455; Gaps 56;			
Qy	4	MTLFFCLLTLSIGWMAQNRVTKGTWISSDNEPLICANVVVGNTTIGAATDLDGNFTL	63
Db	946	MAISALNLSGQW-IAKNLTKANSLTS-----AGDITGVDTLITVNTQL	991
Qy	64	SVPANAKMLRVSGMTTKEVAIANVMKIVLDPDSKYLEQVVLGYGTGQKLS-TVSGSV	122

Db	992	NNQANGKLL--SAGVLTALKADSVTNDGQLQGNVTTITAGQLTNGHGLQGETLTLTASGV	1049
Qy	123	AKVSSKLAEPKPVANTMDA-----LQGOVAGMQVMTT-----SGDPTAVASVE	165
Db	1050	NRRSGGVLSRNALNVSTATLSNQSTIQGG-GGVLSNATDRLQNDQKILSGSNLTITAOV	1108
Qy	166	IHGTSLSGASSAPLYIVDGMQT-----SLDYVAT-----MNPNDPE	201
Db	1109	LANTGS-GLVQAATLLLDVVNTVNGGRVLATGSADVKGKTLNNTGTLOGAXLVNVHTFS	1167
Qy	202	SMSVLKADASATSIYGARAANGVVIQTKKGJMSERGITFNAS--YGISQILNTKPLDNM	259
Db	1168	XSGTLLGTSLGVKGSS-----LLQNGTGRLYSAGNLLLDADQDFSGQGV-----V	1213
Qy	260	MTCD-----ELLDFQVKAGFWGNQTVQVK-----DMILAGAEOLYGNVSLKD	304
Db	1214	ATGDVTLKIAALTNHGTLAAGKTLVSQNAITNGVWGQDAMVLOGABFTNKGXLT	1273
Qy	305	EYKTLFPVDFNHDADWLKALFKTAP-----TSQGDISFSG--GSGQTSYASI	351
Db	1274	GKNSVPSA-----QRLFLNAPGSLQGGDVSLSNRSRSDITISGFTGTAGSLTMNVA	1324
Qy	352	GYPDQEGMAREPANFKRYSGRLNPE-----SRINELKV-----GANL-----	389
Db	1325	GTLNLSALIVAGNNKLFTDRLHNOHGDILAGNSLWQKDSAGGANTTEINNSGNIETHQ	1384
Qy	390	-----SGAIANPRRSADYFGKYMGSGTFCVLTMPRYNPFVNG--DLADV-----	433
Db	1385	GDIVVTRTHLLNQREG-----FSATTTTR-TNPSSIQGMNALVDIPLSLLPD	1431
Qy	434	-----YYM-----YGATRPSMTEPYPAKRPFSSESHQANVNGFAQITPKGL	476
Db	1432	GSVGYTFREVENOHGTPCNGHGGACNITMDTLYY--YAPFADSATORFLSS-QNITVTGA	1488
Qy	477	TLKAQAGVITNTRTSSKMPNPNYDSTPLGERRERAYRDVSKFTNTABYKESIDEKHD	536
Db	1489	DNPAGRIASGRNLSAEERLENRASFILANG-----DIALSGRELSNQSQGTGTE--	1538
Qy	537	LTALMGHEVIEYEGD--VIGASSKGFESDKMLLSQKGTGNSLSLPEHRAEYAVLSFF	593
Db	1539	-----NEVLVRYDPKTFYGSYATG-SLDKGLPLSPENNTI-----	1575
Qy	594	SRENYGDKWMIYDFSVRNDQSRFGSNR-----SAMPYSV--GGMFDIYNKFIQES	644
Db	1576	-----RFLSDGREKDYTPGKTYYSVIQAGG--DVKTRFTS--	1608
Qy	645	NWLSDLRLKMSYGTGTSNBIQNNYHOALVTNNYTEDAMGLSISTAGNPDLSWEKOSQFN	704
Db	1609	-----SINNGTTTAHAGSVSP-----	1624
Qy	705	PGLAAGAFNNRLSAEVDYVVRTTNDMLIDVPMPIYISGFFSOYQNV-----GSM	752
Db	1625	--VVSAPVLNLSQ--TGDSLTOTAL-----QQEPVVVGVSPQWDELALAGAL	1669
Qy	753	KNTGVDSLKGTIYQNKDNWVYASANFNYNRQBITKLFFGLNKYMLPNTGTIWEIGYPSN	812
Db	1670	KNIAGCSPLTGTGTSIDDW-----PLPSGNGYLVPSD-----PDS	1706
Qy	813	FYMAE-----YAGIDKKTGQLWYVPGQVDADGNKVTTSQYSADLETRID	857
Db	1707	PYLITVNPKLGLQVDSHLFAGLYELLGAK-----PCQAPRE-----TAPSYTDEKQ	1754
Qy	858	KSVTPPTITGGFSGISGASW-----KGLSLDADPFIY-----VGKWMNLN---	893
Db	1755	-----FUGSSYFLDRGLGKPEKDYRFUGDAVDFDTRYVSVNAVLSRTGSRYNGLG	1803
Qy	894	-----NDRYFTENAG-----GLMQLNKDKMLLNNAWTEONKETDVPKLGQ	932
Db	1804	SDTEQRYLMDNAAARQKGLGLEFVALTAEQIADQLDGSILWWSVTINQOTVMPKLYL	1863
Qy	933	SPOFDTHLLENASFLRLKMLKLYVLPNLSFAGQNVII--GGARVYLMARNLLTIVTKYKGF	990
Db	1864	SPEDIT--LHNGSVISGNVQL-----AGGNITNSGGS---INAQNDLSLDSSGYI	1909

QY 991 DP-----EAGNV 998
| | | | |
Db 1910 DNLNAGLISAGGSL 1923

RESULT 11
US-09-895-913A-12
; Sequence 12, Application US/09895913A
; Patent No. US20020160456A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020160456A1 Helicobacter Polypeptides in the
; FILE REFERENCE: 06132/043002
; CURRENT APPLICATION NUMBER: US/09/895,913A
; PRIOR FILING DATE: 2001-06-29
; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 1974
; TYPE: PRF
; ORGANISM: Helicobacter pylori
US-09-895-913A-12

Query Match 3.2%; Score 167; DB 10; Length 1974;
Best Local Similarity 19.0%; Pred. No. 0.00048;
Matches 243; Conservative 155; Mismatches 456; Indels 428; Gaps 59;

QY 13 TSIGWMAQRVTKGVTSSENEPL-----IGANVV-----V 46
| | | | |
Db 110 SAFWLSNFGATVGTLLGSAQKANNNGSIWFGKNLLYLHGNFATNIFLTNNFV 169
| | | | |
QY 47 GNTTGAATDLDGNFTLSPANAKMLRVSYSGMTTKEVAIANVMKVLDPDSKVLQVVV 106
| | | | |
Db 170 GNPAGGGGAT--NFNAEDTLNADGL--NYTF-----QTV 202
| | | | |
QY 107 LGYGTGQKLTSGSVKAVSSKLAEPKVPANIMDALQGVAGQVMTTSGDPTAVASVEI 166
| | | | |
Db 203 LGLQT-----SASQHSWANFNKLSMEIKNSFRDFTWGGFNFGRIPTFTFGWTNI 258
| | | | |
QY 167 HGTGSLGAS-----SAPLYVDGMQTSLDVATMN----- 196
| | | | |
Db 259 NGATESGSSYVNVNVAANTDLIFNSILGGGIRYDLKANNIIFNNSQWVIDVSKNVQSSLN 318
| | | | |
QY 197 -----PN-----DPESNVLKASATSIY-----GARAANGVFIQTK 229
| | | | |
Db 319 GNVTFNNSLSVKPNAAINIGDSQTALENASSLSFYNNVANFNGTTFAGVSYLNLN 378
| | | | |
QY 230 KGKMSGRITFNAS-----YGISQLNTKPLONMTGDELLQVQKAGFGWGNQTVQVK 285
| | | | |
Db 379 PNAQVSFNVQVNFNNAVTFYGLPGKTPDFGN---SARLINFKNQTNF---NQATLNR 432
| | | | |
QY 286 -----DMTLAGAEDLYGNYDSLKDEYKTLF----- 311
| | | | |
Db 433 AKNIHINFQGVTFKQNSTWMLAESQASFNALKVE-GETNFNLNSSLNFGNSVFNA 491
| | | | |
QY 312 PVPD--NHADWLKALFKTAPTISOQDISFGSGSQGTSYASIGYFDQEGMAREPANFKRY 369
| | | | |
Db 492 PVSFYANHS-----QISFTKLATFNSDASFDLSNNST----- 523
| | | | |
QY 370 SGRLPFESRINELWLVKANLGAIANRRSADYFGKYMGSGTFFGLVTPRYNPFVDVNGD 429
| | | | |
Db 524 ----LNFQSVL---LNGALNLLGNGNNLAINAKGNFSGSK--GIILNL--SYNMLF--GSD 572
| | | | |
QY 430 -LADVYMYGATRPS--MTEPYFAKMRPFSSSHQAN-----VNGFAQITPIKGL----- 476
| | | | |

Db 573 KTSYDVLOAQNIIDGLMNGNGYKIRFYGIQIDRADYSFONGVHSMRFTNPLNTTETIT 632
| | | | |
QY 477 -TL-KAAGAVDITNTRTSSKMPN-----NPYDSTPLGERREYAYR---DVSKSF 521
| | | | |
Db 633 ETLHNRLKVLISQNGVSNKMFNLAPSLYDYQKNPYNET-----ENSYNTSKVGY 686
| | | | |
QY 522 TNAEYK-FSIDEK-----HDLTALMGHEY-----IEYBGDVIGASSKGFESDKLML 567
| | | | |
Db 687 YLTSNIKGFNQNNKTPGTYNACNQPLQAL--HIYNOAITKQDLNMIASIGKEFLPKIANL 744
| | | | |
QY 568 LSQ-----KTCNSLSLPEHRVAEYAYLSPSRFNYGDKMYIDF 608
| | | | |
Db 745 LSSGALDNLNPNSEFETLFGIFEKYGITLQENWKSLLKIINNFSNTTNYDFSQGNLVVG 804
| | | | |
QY 609 SVRNDQSSRFSGNNRS-AWFYSVG-----GMFDIYNKFIQESN-----WL--S 648
| | | | |
Db 805 AIKEQ-----TNTKSVVWFGGEGYKPCAVGDNTQCMFRQTNLQQLHSSSTPYLGYN 859
| | | | |
QY 649 DLRLKMSY--GTTGNSEIGNYNHQAIVTVNNYTE-----DAMGLS-----ISTAGNP 693
| | | | |
Db 860 NFRAKNIYITGTIGSGNAWGGSGSANVSFESCTNLVLNQAIDAQGTDKIFSYLQGGIE 919
| | | | |
QY 694 DLSWEKQSQFNFGLAAGAFNNRLSAEVDYFVTTNDMLID-----VMPY-----ISGPF 743
| | | | |
Db 920 KLFGEK-----GLGNALSNI--YEESLNDNAIPKDLANMIKDFGSKTLLSLL 966
| | | | |
QY 744 SQYQ-----NVGSMKNTGVDLSLKTIIYQNKDMNVYASANFNYNRQETIKLFF----- 791
| | | | |
Db 967 SPTEVNLLGVSAFKAIMEILNSKTVGDVFGENGLLNALDPTERRKIDQMLLEQIQAH 1026
| | | | |
QY 792 -GLNKYMLPNTGTIWEIGYPNSFYMAEYAGIDKKTGKQLW-----YVPGQVDADGNKVT 844
| | | | |
Db 1027 SGFEKFIYKTLGIENVENFINWY-----GKQSLSPANNFVPGGLNQALDKIG 1075
| | | | |
QY 845 TSQVSADLETIDKSVTPITGGSLSGA-----SWKG-----LSLDADPAY----- 885
| | | | |
Db 1076 SSSDAKDLQNFDKTFTFGDILNQMIQEAFLINKLISWLGPDLSVLVIALNSITNPSKE 1135
| | | | |
QY 886 -----IVGKWMIN-----NDRYFTENAGGLMQLNKKMLLN 918
| | | | |
Db 1136 LTSTISSIGEALNDLLDGGVNVKIMSQVLMQINKIIADKPGGVYQQGLGSL----- 1191
| | | | |
QY 919 TEDNKETDVPKLGOSQPDTHLLENASFLRLNKLKLYVLNPSLPAGQ-----N 967
| | | | |
Db 1192 -PQSLQDELKKGSLGSLGSRGLNL-----WQGNFNFVAKDYLFNTNSSFSGNATGGLN 1246
| | | | |
QY 968 VIGGARVYLMARNLLTVTKYKG 989
| | | | |
Db 1247 FVAGKSIIFNGKNTINFQYQG 1268
| | | | |

RESULT 12

US-10-238-075-1119
; Sequence 1119, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated
; FILE OF INVENTION: E.coli, and biological uses of these polynucleotides and of their
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1119
; LENGTH: 2732
; TYPE: PRF
; ORGANISM: Escherichia coli
US-10-238-075-1119


```
QY 400 ----DYFGKYMGSGTGVLT-----PRYY--NPFVNDLADVYMYGA-- 439
Db 817 ALNGLFSDGHIQAGKSKITLSTGPKVKTANOAPAVYLTGVDLTGDNATLEITRGHGA 876
QY 440 -----TRPSWTEPYFAKMPFSS-----ESHQANVGF-----AQITPIKG 475
Db 877 SGOIHASASTVIGSDTTPAELASAEATTASAFAGSLLEGYNAAFNAITGGRADVSMHNA 936
QY 476 L-TLKAQAGV---DITNTRTSSKRMNNPYDSTPLGERRERAYRDVSKFTTAYKFSI 531
Db 937 LWTLGDSALHTLTVNSRISB-----GDRIFRTL-TVKNKLATGSDPEVLR 983
QY 532 DEKHDLTALMGHIEYIEYEDGVICASSKGFESDKMLLSQK---TGNLSLSLPEHRVAEYA 588
Db 984 DLKN-----ADKINVTEKATGSDNLSNVFPMKOPAQOQSLNIP--LVTAPA 1027
QY 589 YLSF-----PSRF-----NYGFDKMYIDFSVRNDQSSRFGSNNRNSAFYS 629
Db 1028 GTSAPFKAGTRMIGFSRVPTLHVDTSGENTKWLIDGPKAEADKAAAKADS----FMN 1083
QY 630 VGMFDIYNKFIQESNMLS-----DLRLKMSYGTGNSEIGNYNHQAALVTNNVTEADML 685
Db 1084 AG-----YKNFMEVNNLANKMGDLR-----DTNGDA-----GAWAR 1115
QY 686 SISTAGNPOLSEKQSQFNLGAAGAFNNRLSAEVDP---YVRTNDMLIDVPMPIYISGF 742
Db 1116 IMSGASAD-----GGYSDNVTHVQVGFDDKXHELDGVDLFTGVTWTVTDS 1161
QY 743 FSOYQNVGSMKNTGVDLSLKGITIQKNDNVYASANFNVRQEIYKFLGKMKMLPNTG 802
Db 1162 ADHAFSGTKSVGGGL-----YASALF-----ESGAYIDLIGKYI----- 1197
QY 803 TIWEIGYPNFYMAYEAGI-DKKTGKQLWYVPGQVADGKNTVTSQYSADLETRIDKSVT 861
Db 1198 -----HNDNDYTGNGAGLTGKHYNTHSWYAGAE-----TGRYHLTEFTFEPQA- 1242
QY 862 PPTGTFSGFSGAS--WKGSLD-----ADPAYIVGK-----W- 890
Db 1243 -ELVYGAVGSKTFRWKGDMGLSMKNRDFSPILGRGTIELGKTFSGKDWVSTARAGTSWQ 1301
QY 891 --MINDRVFTENAGLMQLNKDK---ML-----LNAWTEDN 922
Db 1302 FOLLNNGETVLRDASGEKRIKGEKSRMLFNVGMNAQIKDN 1342

RESULT 14
US-09-882-227-522.
; Sequence 522, Application US/09882227
; Publication No. US20030158396A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Ooomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20030158396A1el Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/047002
; CURRENT APPLICATION NUMBER: US/09/882,227
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/902,615
; PRIOR FILING DATE: 1997-07-29
; NUMBER OF SEQ ID NOS: 638
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 522
; LENGTH: 2893
; TYPE: prt
; ORGANISM: Helicobacter pylori
US-09-882-227-522
```

Query Match 3.0%; Score 160; DB 12; Length 2893;
Best Local Similarity 19.3%; Pred. No. 0.0034;

```
Matches 227; Conservative 135; Mismatches 399; Indels 414; Gaps 59;
QY 39 IGANVVV-VGNTTICAAATDLDGNFTLSYPANAKMLRVSYSGWTT-KEVAIANVMKIVLDP 96
Db 134 VGGNITINLGSVV---LDLSGSNSFT-----SYQYNOGKDDVFTTVGAINLNG 170
QY 97 DSKVLEQVVVLGYGTGQKLSLTVSGSV-----AKVSEKLAEPKXANIMDALOGQVAG 148
Db 171 TLEVGNRV---GSGAGTGTATLNLANKVNINSINAYKTSQVNICN-----ANSVIT 222
QY 149 MQVMTTSDP-TAVASVEIHGTSLGASSAPLYIVDGMQTSLDVVDVATWNPNDPESMSVLK 207
Db 223 IGSVLSLSDGVDCSSLASV---GIGANCSTSGPSYFKG----- 256
QY 208 DASATSIYGARAAANGVWFIO--TKGKMSERGRITFNASY-----GISO 249
Db 257 TTNATNATFASNASGSGFTFEENATPSGAKWNGGTYTFNKEFATNTATFSSSGSNFKGVSS 316
QY 250 ILNTKPLDNMTGDELLDFQVKAGFWG-----NNQT-----VQKVKDMILAGAE 294
Db 317 FNGTSFNSASVTFDNOATFQ--NSSFNGGCTFTFNQTNPTNNAQHPQIQN-----SS 366
QY 295 LYGNVDSLKDEYKTLFPVDF---NHDADWLKALFKTAP-----TSQDISFSGSGSQ 344
Db 367 FSGNATTLK---GFYNFOQAFNNNSHQLTIQNASFNNTATFNNTGKITTEKDAFNTTTFN 423
QY 345 TSYIASIGVPEQEGMAREPANFKRYSGRLNFESRINEWLKVGANLSGA--TANRESADY- 401
Db 424 TS-----VDTNMS-----VTGGVTLUSGKNLKNGSTLDFG 454
QY 402 -----FGKYMGSG-----GTGVLTPRYNPFVNDGDLADVYMYGATRPS 443
Db 455 SSKITLAGTTFNLTSLGSEKSVTILNSSGGITYSNLNLN-HAINGLTSAI-----KTWES 508
QY 444 MTEPYFAKMPFSSSHQANVNGFAQITPIKGLTLKAQGVDTITNRTSSKMPNPNVDS 503
Db 509 LSNP-----QSPAQLMDIITYNGVT-----GOLLNENAAATSK-----PTDS 545
QY 504 TPLGERRERAYRDSKFTNTAE-----YKFSIDKHDLTALMGHEYIEY--- 548
Db 546 SP-----SKSSTNSTQVYQVGYKIGDTIYKLOETFESHNSIIIIQALLESYTPP 593
QY 549 -----EGDVIGASSKGFSDKMLLSQGT-GNSLSLPEHRVAEYAYLSPFSRF 596
Db 594 PVINGSKFDLSASNYINADMPWDHKYIIPKSQNTFESGTYVLPSPVQVWGSYTNFKQTF 653
QY 597 N-----YGFDKMYIDFSVRNDQSSRFGSNNRSA-----WPFY-----SVGMFDIY 637
Db 654 SANGSNLVIGYNS-TWTDHNVSSSGTSGVSGDTSALNGHCGPWPYQCTGTGTNGTYSAY 712
QY 638 NKFIQESNWLSDLRLKMSYGTG-----NSEIGNYN---HOALVTNNVT 679
Db 713 HVYI-----TANLRSNGRIGTGGAANLIFNGVDSINIANATITQHNAGIYSSMTFTSQS 767
QY 680 ED-----AMGLS-----ISTAGNPDLSEKQSQFNLGAAGAFNNRLSAEVDVYRTNDML 731
Db 768 MDNSQNLNGLNSGKLSVYGTFTTNEAKDGFIFNAGQAVFEN-----TN--- 812
QY 732 IDVPMPIYSGPFSQVNVGSMKNTGVDLSLKGITIQKNDMNVAANFNYNRQETIKULFF 791
Db 813 FNGGSYQPSGDSLNFNNNQFNNSGSGFEISAKNASFNNAFN--NSASFNFNNSNATTSFV 870
QY 792 G-----LNKYMPLPNTCTIWEIGYPNFYSFMAEYAGIDKKTGK 827
Db 871 GDTNANSNLQIAGNAVFNSTNGSQNTANFNNTGSV-NISGNATFDVNVNFGPTNTS-- 927
QY 828 QLWYVPGQVDADGNKVTTSQYSADLE-----TRIDKSVT--PPII----- 865
Db 928 ----VKQVTL--NNITLKNLNAPLSFGDGTITFNAHSVINIAESITNGNITLVSSSKE 981
QY 866 ----GGFSL-----GASWGLSLDA-----DPAYIVGKWMINNDRYFTENAGGL 905
Db 982 TEYNNAFSKNLWQLINYQCHGASSEKLVSSAGNGYDVVYVFNQTYNFQVFSQNS--- 1038
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2004, 19:04:23 ; Search time 47 Seconds
(without alignments)
3434.573 Million cell up

Title: US-09-581-286A-424
 Perfect score: 1017
 Sequence: 1 MKRMTLFFCLLLTSIGWAMA.....VGRNOYPNSKQYVAGIOLSF 1017

Scoring table: OLIGO
Gapop 60.0 . Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0 No word size was specified. Sorry. *Do you need it now?*
Total number of hits satisfying chosen parameters: 1107863 *Let me know.*

Total number of hits satisfying chosen parameters: 1107863

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Listing first 45 summaries

Database : A Geneseq 19Jun03:*

1:	SIDS1	gcgdata	/geneseq	/geneseq	emb1	/AA1980	DAT.*
2:	SIDS1	gcgdata	/geneseq	/geneseq	emb1	/AA1981	DAT.*
3:	SIDS1	gcgdata	/geneseq	/geneseq	emb1	/AA1982	DAT.*
4:	SIDS1	gcgdata	/geneseq	/geneseq	emb1	/AA1983	DAT.*
5:	SIDS1	gcgdata	/geneseq	/geneseq	emb1	/AA1984	DAT.*
6:	SIDS1	gcgdata	/geneseq	/geneseq	emb1	/AA1985	DAT.*
7:	SIDS1	gcgdata	/geneseq	/geneseq	emb1	/AA1986	DAT.*
8:	SIDS1	gcgdata	/geneseq	/geneseq	emb1	/AA1987	DAT.*
9:	SIDS1	gcgdata	/geneseq	/geneseq	emb1	/AA1988	DAT.*
10:	SIDS1	gcgdata	/geneseq	/geneseq	emb1	/AA1989	DAT.*
11:	SIDS1	gcgdata	/geneseq	/geneseq	emb1	/AA1990	DAT.*
12:	SIDS1	gcgdata	/geneseq	/geneseq	emb1	/AA1991	DAT.*
13:	SIDS1	gcgdata	/geneseq	/geneseq	emb1	/AA1992	DAT.*
14:	SIDS1	gcgdata	/geneseq	/geneseq	emb1	/AA1993	DAT.*
15:	SIDS1	gcgdata	/geneseq	/geneseq	emb1	/AA1994	DAT.*
16:	SIDS1	gcgdata	/geneseq	/geneseq	emb1	/AA1995	DAT.*
17:	SIDS1	gcgdata	/geneseq	/geneseq	emb1	/AA1996	DAT.*
18:	SIDS1	gcgdata	/geneseq	/geneseq	emb1	/AA1997	DAT.*
19:	SIDS1	gcgdata	/geneseq	/geneseq	emb1	/AA1998	DAT.*
20:	SIDS1	gcgdata	/geneseq	/geneseq	emb1	/AA1999	DAT.*
21:	SIDS1	gcgdata	/geneseq	/geneseq	emb1	/AA2000	DAT.*
22:	SIDS1	gcgdata	/geneseq	/geneseq	emb1	/AA2001	DAT.*
23:	SIDS1	gcgdata	/geneseq	/geneseq	emb1	/AA2002	DAT.*
24:	SIDS1	gcgdata	/geneseq	/geneseq	emb1	/AA2003	DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	1017	100.0	1017	20	AAV344477	Porphyromonas ging
2	1017	100.0	1046	20	AAV343453	Porphyromonas ging
3	913	89.8	1014	20	AAV344478	Porphyromonas ging
4	9	0.9	100	22	AAV86385	Human immature/haema
5	8	0.8	239	24	ABP71236	M. catarrhalis Mid
6	8	0.8	343	22	AAU34481	E. coli cellular p
7	8	0.8	476	22	ABG08740	Novel human diagno
8	8	0.8	487	24	ABU00872	S. pneumoniae type
9	8	0.8	507	20	AAV34756	C. pneumoniae prot

ALIGNMENTS

RESULT 1

RECOIL I
AAY34477
ID AAY34477 standard: 1017 AA.

AA
AC
AAV34477:

DT	20-MAR-2003	(updated)
DT	25-AUG-1999	(first entry)

XX
DE
Porphorymonas gingivalis protein PG2.

AA Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
KW vaccine: antigenic
KW

XX QS *Pamphorymonas gingivalis*

XX PN WO9929870-A1

XX
PD
17-JTN-1999

XX
PF 10-DEC-1998: 98WQ-A1101023

XX
PR 10-DEC-1997: 97AII-0000839

PR	31-DEC-1997;	97AU-0001182.
PR	30-JAN-1998:	98AU-0001546.

PR 10-MAR-1998; 98AU=0002264.
PR 09-APR-1998; 98AU=0002911

PR	23-APR-1998;	98AU-0003128.
PR	05-MAY-1998:	98AU-0003338.

PR 22-MAY-1998; 98AU-0003654.
PR 29-JUL-1998; 98AU-0004917.

PR 30-JUL-1998; 98AU-0004963.

Aspergillus fumigatus
Amino acid sequence
Aspergillus fumigatus
Herbicide-activated
Moxarella 200 kDa
M. catarhalis str.
M. catarhalis M56
M. catarhalis str.
M. catarhalis str.
Moxarella catarhalis
Moxarella catarhalis
M. catarhalis sur.
M. catarhalis les.
Human gene 39
Human albumin fusi
Propionibacterium
S. pneumoniae dity
Human novel foetal
Swinepox virus H1n1
Arabidopsis thaliana
Sequence of a poly
Sequence of a poly
Human ORFX ORF937
Human synthase-like
Human ORFX protein
Sequence of a poly
Human liver peptid
Peptide #8731 enco
Protein #7230 enco
Human brain expres
Human bone marrow
Peptide #9044 enco
Human peptide enco
Rat metalloprotein
Arabidopsis thaliana
Human ORFX protein

PR 04-AUG-1998; 98AU-0005028.
 PA (CSLC-) CSL LTD.
 XX Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
 PI Ross BC, Rothel LJ, Webb EA;
 XX WPI; 1999-385613/32.
 DR N-PSDB; AAX91695.
 XX
 PT Antigenic Porphyromonas gingivalis peptides for preventing
 PT gingivitis
 XX
 PS Claim 1; Page 456-458; 588pp; English.
 XX
 CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
 CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAX94318 to
 CC AAX94583. AAX91802 to AAX91989 represent PCR primers used in the
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The PG polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis.
 CC (Updated on 20-MAR-2003 to correct PR field.)
 XX
 SQ Sequence 1017 AA;

Query Match 100.0%; Score 1017; DB 20; Length 1017;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1017; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKBMTLFFCLLSIGWMAQNRVTGTVISSENEPLIGANVVVGVNTTIGATDLGDN 60
 DB 1 MKBMTLFFCLLSIGWMAQNRVTGTVISSENEPLIGANVVVGVNTTIGATDLGDN 60
 QY 61 FTLSVPANAKMLRVSYSGMTTKEVAIANVMKIVLDPDPSKVLEQVVLGYGTGKLSVSG 120
 DB 61 FTLSVPANAKMLRVSYSGMTTKEVAIANVMKIVLDPDPSKVLEQVVLGYGTGKLSVSG 120
 QY 121 SVAKVSSEKLAEPVANIMDALQGVAGMOMVTSGDPTAVASVEIHGTGSLGASSAPLY 180
 DB 121 SVAKVSSEKLAEPVANIMDALQGVAGMOMVTSGDPTAVASVEIHGTGSLGASSAPLY 180
 QY 181 IVDMQTSLDVATMNPDPESVLDKASATSYGARANGVVFVQTKGKRSERGIT 240
 DB 181 IVDMQTSLDVATMNPDPESVLDKASATSYGARANGVVFVQTKGKRSERGIT 240
 QY 241 FNASYGISQILNTKPLDNMMTGDELLDFQVKGAFWGNQTVQVKMILAGAEADLYGND 300
 DB 241 FNASYGISQILNTKPLDNMMTGDELLDFQVKGAFWGNQTVQVKMILAGAEADLYGND 300
 QY 301 SLKDEYKTLFPVDFNHDADWLKALFKTAPTSQGDISFSGSGSQSTSYASIGYFDQEGMA 360
 DB 301 SLKDEYKTLFPVDFNHDADWLKALFKTAPTSQGDISFSGSGSQSTSYASIGYFDQEGMA 360
 QY 361 REPANFKYGRNLNFESRINEWLKVGANLSGATANRRSADYFGKYNGSGTFCGLTWPRY 420
 DB 361 REPANFKYGRNLNFESRINEWLKVGANLSGATANRRSADYFGKYNGSGTFCGLTWPRY 420
 QY 421 YNPFVNGDLADVYMYGATRPSTPEYFAKMRPFSSSHQANVNGFAQITPIKGLTLKA 480
 DB 421 YNPFVNGDLADVYMYGATRPSTPEYFAKMRPFSSSHQANVNGFAQITPIKGLTLKA 480
 QY 481 QAGVDITNTRTSKRMNNPYDSTPLGERRERAYRDSKFTNTAEYKFSIDEKHDLTAL 540
 DB 481 QAGVDITNTRTSKRMNNPYDSTPLGERRERAYRDSKFTNTAEYKFSIDEKHDLTAL 540
 QY 541 MGHEYIEYEGDVTGASSKGFSKMLLSQKGTGNSLSLPEHRVAEYAYLSFSSRNYGF 600
 DB 541 MGHEYIEYEGDVTGASSKGFSKMLLSQKGTGNSLSLPEHRVAEYAYLSFSSRNYGF 600
 QY 601 DKWYIDFSVRNDQSSRFSGNNSAWFYSGVGMFDIYNKFQIESNMLSDLRLKMSYGTG 660

DB 601 DKWYIDFSVRNDQSSRFSGNNSAWFYSGVGMFDIYNKFQIESNMLSDLRLKMSYGTG 660
 QY 661 NSEIGNYNHQAALVTNNYTEDAMGLSISTAGNPDLSWEKOSQFNFLAAGAFNNRLSAEV 720
 DB 661 NSEIGNYNHQAALVTNNYTEDAMGLSISTAGNPDLSWEKOSQFNFLAAGAFNNRLSAEV 720
 QY 721 DFYVRTTNDMLIDVPMPIYISGPFSSQYQNVGSMKNTGVDLSLKGTYIQNKDMNYYASANFN 780
 DB 721 DFYVRTTNDMLIDVPMPIYISGPFSSQYQNVGSMKNTGVDLSLKGTYIQNKDMNYYASANFN 780
 QY 781 YNRQETITKLPFGLNKYMMLNTGTIWEIGYPNSFYMAEYAGIDKKTQKQIWWYVPGQVDADG 840
 DB 781 YNRQETITKLPFGLNKYMMLNTGTIWEIGYPNSFYMAEYAGIDKKTQKQIWWYVPGQVDADG 840
 QY 841 NKVTSQYSADLETRIDKSVTPPTITGGFSLGASWKLSDADPFIYVKGWMINNDRYETE 900
 DB 841 NKVTSQYSADLETRIDKSVTPPTITGGFSLGASWKLSDADPFIYVKGWMINNDRYETE 900
 QY 901 NAGGLMQLNKDKMLNNAWTDNKETDVPKLGOSQPDTHLLENASFLRLKNLKLTIVLPN 960
 DB 901 NAGGLMQLNKDKMLNNAWTDNKETDVPKLGOSQPDTHLLENASFLRLKNLKLTIVLPN 960
 QY 961 SLFAGQNVIGGARVYLMARNLLTVTKYKGFDPGAGNVGNQYPSKQYVAGIQLSF 1017
 DB 961 SLFAGQNVIGGARVYLMARNLLTVTKYKGFDPGAGNVGNQYPSKQYVAGIQLSF 1017

RESULT 2
 AAX34353
 ID AAX34353 standard; Protein; 1046 AA.
 XX AAX34353;
 AC
 XX
 DT 20-MAR-2003 (updated)
 DT 25-AUG-1999 (first entry)
 XX
 DE Porphyromonas gingivalis protein PG2.
 XX
 KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
 KW vaccine; antigenic.
 XX
 OS Porphyromonas gingivalis.
 XX
 PN WO9929870-AI.
 XX
 PD 17-JUN-1999.
 XX
 PF 10-DEC-1998; 98WO-AU01023.
 XX
 PF 10-DEC-1997; 97AU-0000839.
 XX
 PR 31-DEC-1997; 97AU-0001182.
 PR 30-JAN-1998; 98AU-0001546.
 PR 10-MAR-1998; 98AU-0002264.
 PR 09-APR-1998; 98AU-0002911.
 PR 23-APR-1998; 98AU-0003128.
 PR 05-MAY-1998; 98AU-0003338.
 PR 22-MAY-1998; 98AU-0003654.
 PR 29-JUL-1998; 98AU-0004917.
 PR 30-JUL-1998; 98AU-0004963.
 PR 04-AUG-1998; 98AU-0005028.
 XX
 XX (CSLC-) CSL LTD.
 XX
 PI Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
 PI Ross BC, Rothel LJ, Webb EA;
 XX
 DR WPI; 1999-385613/32.
 DR N-PSDB; AAX91571.
 XX
 PT Antigenic Porphyromonas gingivalis peptides for preventing
 PT gingivitis
 XX

SQ Sequence 1014 AA;

Query Match 89.8%; Score 913; DB 20; Length 1014;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1013; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	4	MTLFFCLLTISIGWAMAQRRTVKGTVISSDNEPLIGANVVVVGNTTIGAAATDLGNTFL	63
Db	1	MTLFFCLLTISIGWAMAQRRTVKGTVISSDNEPLIGANVVVVGNTTIGAAATDLGNTFL	60
Qy	64	SVPANAKMLRVSYSGMTTKEVAIANVMKIVLDPDSKVLEQVVVLYGYTGQKLSVTSVSSVA	123
Db	61	SVPANAKMLRVSYSGMTTKEVAIANVMKIVLDPDSKVLEQVVVLYGYTGQKLSVTSVSSVA	120
Qy	124	KVSEKLAKEPVANIMDALQOVAGMOMTTSQDPTAVASVEIHGTGSLGASSAPLYITVD	183
Db	121	KVSEKLAKEPVANIMDALQOVAGMOMTTSQDPTAVASVEIHGTGSLGASSAPLYITVD	180
Qy	184	GMQTSLDVATMNPNDPESMSVLKASATSIIYGARAANGVVFIQTKGKMSERGRITFNA	243
Db	181	GMQTSLDVATMNPNDPESMSVLKASATSIIYGARAANGVVFIQTKGKMSERGRITFNA	240
Qy	244	SYGISQILNTKPLDNMTGDELLDFQVKAGFWGNNTQVKVKDMILAGAEADLYGNYDSLK	303
Db	241	SYGISQILNTKPLDNMTGDELLDFQVKAGFWGNNTQVKVKDMILAGAEADLYGNYDSLK	300
Qy	304	DEYKTLFPVDFNHDADWLKALFKTAPTSQGDISFSGSGSQTSYVASTGYVDPQEGMAREP	363
Db	301	DEYKTLFPVDFNHDADWLKALFKTAPTSQGDISFSGSGSQTSYVASTGYVDPQEGMAREP	360
Qy	364	ANFKRYSGRLNFESRINEWLKVGANLSGAIANRRSADYFGKYMGSGTGFVLTMPRYNYP	423
Db	361	ADFKRYSGRLNFESRINEWLKVGANLSGAIANRRSADYFGKYMGSGTGFVLTMPRYNYP	420
Qy	424	PDVNGDLADVYMYGATRPSMTPEYFAKMRPFSSSHOANVNGPAQTPPIKGLTLKQAQ	483
Db	421	PDVNGDLADVYMYGATRPSMTPEYFAKMRPFSSSHOANVNGPAQTPPIKGLTLKQAQ	480
Qy	484	VDINTRTSSKRPNNPVDSYPLGERRERARDYKSFNTAEYKFSIDEKHDLTALMGH	543
Db	481	VDINTRTSSKRPNNPVDSYPLGERRERARDYKSFNTAEYKFSIDEKHDLTALMGH	540
Qy	544	EYIEYEGDVI GASSKGESDKMLLSQKTCNSLSLPEHRVAEYAYLSFFSRFNYGFDKW	603
Db	541	EYIEYEGDVI GASSKGESDKMLLSQKTCNSLSLPEHRVAEYAYLSFFSRFNYGFDKW	600
Qy	604	MYIDFSVRNDQSRFGSNRNSAWFYSGVGMFDIYNKFQESNWLSDRLKMSYGTGNSE	663
Db	601	MYIDFSVRNDQSRFGSNRNSAWFYSGVGMFDIYNKFQESNWLSDRLKMSYGTGNSE	660
Qy	664	IGNYNHQAALVTNNYTEDAMGLSISTAGNPDLSEKQSFNGLAAGAFNNRLSAEVDYF	723
Db	661	IGNYNHQAALVTNNYTEDAMGLSISTAGNPDLSEKQSFNGLAAGAFNNRLSAEVDYF	720
Qy	724	VRTNDMLIDVPMPYISGFFSQYQNVGSMKNTGVDLSLKGTYIQNKDNNVYASANFNYNR	783
Db	721	VRTNDMLIDVPMPYISGFFSQYQNVGSMKNTGVDLSLKGTYIQNKDNNVYASANFNYNR	780
Qy	784	QEIITKLPFLGNKMYL PNTGTIWEIGYPNFSFYMAEYAGIDKKTGKOLWYVPQVDADGNKV	843
Db	781	QEIITKLPFLGNKMYL PNTGTIWEIGYPNFSFYMAEYAGIDKKTGKOLWYVPQVDADGNKV	840
Qy	844	TTTSQYADLETRIDKSVTPPTITGFSGLASWKGLSLDADFAIVGKMNNDRYFTENAG	903
Db	841	TTTSQYADLETRIDKSVTPPTITGFSGLASWKGLSLDADFAIVGKMNNDRYFTENAG	900
Qy	904	GLMQLNKDKMLLNATWEDNKETDVPKLGQSQQFQDTHLLENASFLRLKNLKLTYVLPNSLF	963
Db	901	GLMQLNKDKMLLNATWEDNKETDVPKLGQSQQFQDTHLLENASFLRLKNLKLTYVLPNSLF	960
Qy	964	AGQNVIGARVYLMARNLLTVTYKGFDPPEAGGNGVGNQYKPNQYVAGIQLSF	1017
Db	961	AGQNVIGARVYLMARNLLTVTYKGFDPPEAGGNGVGNQYKPNQYVAGIQLSF	1014

RESULT 4

AAM86385
ID AAM86385 standard; Protein; 100 AA.

XX AAM86385;

XX AC
XX DT 07-NOV-2001 (first entry)

XX DE Human immune/haematopoietic antigen SEQ ID NO:13978.

XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX KW cytostatic; gene therapy; vaccine; metastasis.

XX OS Homo sapiens.

XX PN WO200157182-A2.

XX PD 09-AUG-2001.

XX PP 17-JAN-2001; 2001WO-US01354.

XX PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 14-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 18-AUG-2000; 2000US-0225759.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.

PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.


```
PR 08-SEP-2000; 2000US-0233081.
PR 12-SEP-2000; 2000US-02331968.
PR 14-SEP-2000; 2000US-02332397.
PR 14-SEP-2000; 2000US-02332398.
PR 14-SEP-2000; 2000US-02332399.
PR 14-SEP-2000; 2000US-02332400.
PR 14-SEP-2000; 2000US-0233401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0234584.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.

PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0256719.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX N-PSDB; AAK59166.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
XX
XX Claim 11; SEQ ID NO 13978; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAK62170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/haematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/haematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention.
XX
XX SQ Sequence 100 AA;
XX
XX Query Match 0.9%; Score 9; DB 22; Length 100;
XX Best Local Similarity 100.0%; Pred. No. 1.5;
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 226 IQTKKGKMS 234
XX | | | | |
XX Db 2 IQTKKGKMS 10
XX
XX RESULT 5
XX ABP71296
XX ID ABP71296 standard; Protein; 239 AA.
XX
XX AC ABP71296;
XX
XX DT 28-APR-2003 (first entry)
XX
XX DE M. catarrhalis Mtd protein immunogenic fragment MID962-1200.
XX
XX KW Surface exposed protein; bacterium; immunoglobulin; IgD; immunomodulator;
XX gene therapy; vaccine; mid.
XX
XX OS Moraxella catarrhalis.
XX
XX PN WO2003004651-A1.
XX
XX
```

```

PD 16-JAN-2003.
XX
XX
XX 01-JUL-2002; 2002WO-SE01299.
XX
XX 04-JUL-2001; 2001SE-0002410.
XX
XX (FORS/) FORSGREN A.
XX
XX Forsgren A, Riesebeck K, Janson H;
XX
XX WPI; 2003-221598/21.
XX N-PSDB; ABZ58975.
XX
XX New surface exposed immunoglobulin D-binding protein from Moraxella
XX catarrhalis, useful for treating an autoimmune disease or as vaccine,
XX comprises a molecular weight of 200 kDa -
XX
XX Claim 3; Page 93-94; 98pp; English.
XX
XX The invention relates to a surface exposed immunoglobulin D-binding
XX protein detected in Moraxella catarrhalis. The protein has an apparent
XX molecular weight of 200 kDa, and can selectively bind membrane bound or
XX soluble IgD. The protein or its variant or fragment, is useful in
XX treating an autoimmune disease or as vaccine. The present sequence
XX represents a M. catarrhalis Mid protein immunogenic fragment.
XX
XX
XX Sequence 239 AA;
XX
XX Query Match 0.8%; Score 8; DB 24; Length 239;
XX Best Local Similarity 100.0%; Pred. No. 37;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 836 VDADGNKV 843
DB 196 VDADGNKV 203
|||||
|

RESULT 6
AAU34881
ID AAU34881 standard; Protein; 343 AA.
XX
XX AAU34881;
XX
XX 14-FEB-2002 (first entry)
XX
XX E. coli cellular proliferation protein #462.
XX
XX Antisense; prokaryotic cellular proliferation protein;
XX antibiotic; antibacterial; drug design.
XX
XX Escherichia coli.
XX
XX WO200170955-A2.
XX
XX 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
XX 23-MAY-2000; 2000US-206848P.
XX 26-MAY-2000; 2000US-207727P.
XX 23-OCT-2000; 2000US-242578P.
XX 27-NOV-2000; 2000US-253625P.
XX 22-DEC-2000; 2000US-257931P.
XX 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KU, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX WPI; 2001-611495/70.
XX N-PSDB; AAS52740.
XX

```

```

XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 10474; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX invention is also useful for the identification of potential new targets
XX for antibiotic development. The antisense nucleic acids can also be used
XX to identify proteins used in proliferation, to express these proteins,
XX and to obtain antibodies capable of binding to the expressed proteins.
XX The proteins can be used to screen compounds in rational drug discovery
XX programmes. The antisense nucleic acid sequence is also useful to screen
XX for homologous nucleic acids which are required for cell proliferation in
XX essential prokaryotic cellular proliferation protein.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 343 AA;
XX
XX Query Match 0.8%; Score 8; DB 22; Length 343;
XX Best Local Similarity 100.0%; Pred. No. 52;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 183 DGMQTSID 190
DB 273 DGMQTSID 280
|||||
|

RESULT 7
ABG08740
ID ABG08740 standard; Protein; 476 AA.
XX
XX ABG08740;
XX
XX 13-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #8731.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX N-PSDB; AAS72927.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX

```

PS Claim 20; SEQ ID No 39099; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome CC and gene mapping, and in recombinant production of (II). The CC polynucleotides are also used in diagnostics as expressed sequence tags CC for identifying expressed genes. (I) is useful in gene therapy techniques CC to restore normal activity of (II) or to treat disease states involving CC (II). (II) is useful for generating antibodies against it, detecting or CC quantitating a polypeptide in tissue, as molecular weight markers and as CC a food supplement. (II) and its binding partners are useful in medical CC imaging of sites expressing (II). (I) and (II) are useful for treating CC disorders involving aberrant protein expression or biological activity. CC The polypeptide and polynucleotide sequences have applications in CC diagnostics, forensics, gene mapping, identification of mutations CC responsible for genetic disorders or other traits to assess biodiversity CC and to produce other types of data and products dependent on DNA and CC amino acid sequences. ABG0010-ABG30377 represent novel human CC diagnostic amino acid sequences of the invention. CC Note: The sequence data for this patent did not appear in the printed CC specification, but was obtained in electronic format directly from WIPO CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 476 AA;

Query Match 0.8%; Score 8; DB 22; Length 476;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 PRTSQGDIS 337
|||||||
DB 5 PRTSQGDIS 12

RESULT 8
ABU00872
ID ABU00872 standard; Protein; 487 AA.
AC ABU00872;
DT 11-FEB-2003 (first entry)
XX S. pneumoniae type 4 strain protein from coding region #440.
XX Bacterial meningitis; pneumonia; sepsis; otitis media;
KW ear infection; antiinflammatory; antibacterial; immunostimulant;
KW auditory; respiratory; gene therapy; vaccine.
XX Streptococcus pneumoniae type 4 strain.
OS WO200277021-A2.
PN 03-OCT-2002.
XX 27-MAR-2002; 2002WO-IB02163.
XX 27-MAR-2001; 2001GB-0007658.
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX Masignani V, Tettelin H, Fraser C;
XX WPI; 2003-040579/03.
DR N-PSDB; ABX06152.
XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT useful as medicaments for treating or preventing a disease or infection
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
PT or ear infection
XX Claim 1; SEQ ID No 880; 56pp; English.

XX The invention relates to a protein comprising or having at least 50% identity to any of the 2469 amino acid sequences, identified in the specification (available on a computer readable format), or its fragment, expressed from 2469 of 2489 identified DNA coding regions from the Streptococcus pneumoniae type 4 strain genomic sequence appearing as AB566454. Also included are an antibody which binds one of the proteins, treating a patient by administering the protein, DNA or antibody (in a composition), a kit comprising first and second primers, which are the nucleic acid cited above or fragments between nucleotides 8-100 of a sequence not defined in the specification, for amplifying a target sequence contained within a Streptococcus nucleic acid sequence, where the first primer is substantially complementary to the target sequence and the second primer is substantially complementary to the complement of the target sequence, and where the parts of the primers having substantial complementarity define the termini of the target sequence to be amplified, assay comprising contacting a test compound with the protein, and determining whether the test compound binds to the protein and a Streptococcus pneumoniae bacterium, where one or more genes encoding the proteins has been rendered inactive. The proteins, nucleic acid molecules, antibody and compositions are useful as medicaments for treating or preventing a disease or infection due to streptococcus bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis media or ear infection. They are also useful in developing vaccines, diagnostics and antibiotics. The methods are useful for identifying immunodominant proteins. The present sequence is one of the 2469 proteins expressed by the identified coding regions from the genomic sequence.
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 487 AA;

Query Match 0.8%; Score 8; DB 24; Length 487;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 261 TGDLELDF 268
|||||||
DB 77 TGDLELDF 84

RESULT 9
AAY34756
ID AAY34756 standard; Protein; 507 AA.
XX AAY34756;
AC AAY34756;
DT 13-SEP-1999 (first entry)
XX C. pneumoniae protein involved in metabolism of amino acids.
DE Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope.
XX Chlamydia pneumoniae.
OS WO9927105-A2.
PN 03-JUN-1999.
PD 20-NOV-1998; 98WO-IB01890.
XX 04-NOV-1998; 98US-0107078.
PR 21-NOV-1997; 97FR-0014673.
XX (GEST) GENSET.
PA Griffais R;
XX

DR WPI; 1999-357842/30.
 XX Genome sequence of Chlamydia pneumoniae
 PS Page 744-745; Disclosure; 1912pp; English.
 XX
 CC AAY34584-Y35879 represent the proteins encoded by all the open reading
 CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
 CC C. pneumoniae causes respiratory disease such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
 CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising
 CC epitope of C. pneumoniae.
 XX
 SQ Sequence 507 AA;
 Query Match 0.8%; Score 8; DB 20; Length 507;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 105 VVLGYGTG 112
 Db 325 VVLGYGTG 332
 |||||
 RESULT 10
 ABJ25726
 ID ABJ25726 standard; Protein; 546 AA.
 XX
 AC ABJ25726;
 XX
 XX
 DT 16-APR-2003 (first entry)
 DE Aspergillus fumigatus essential gene protein #384.
 DE
 DE Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
 KW cancer; contamination; biofilm; antibody; immune response.
 KW
 XX Aspergillus fumigatus.
 XX
 XX W0200286090-A2.
 XX
 XX 31-OCT-2002.
 XX
 XX 23-APR-2002; 2002WO-US13142.
 XX
 XX 23-APR-2001; 2001US-285697P.
 PR 27-APR-2001; 2001US-287066P.
 PR 05-JUN-2001; 2001US-295890P.
 PR 09-JUL-2001; 2001US-303899P.
 PR 31-AUG-2001; 2001US-316362P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 XX
 PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
 XX WPI; 2003-093124/08.
 XX
 XX New purified or isolated nucleic acids of essential genes of
 PT Aspergillus fumigatus, useful for treating or preventing infections by
 PT A. fumigatus, or for treating a non-infectious disease in a subject
 PT e.g. cancer
 XX
 XX Disclosure; Page -; 175pp; English.
 PS
 XX The invention relates to novel purified or isolated nucleic acids of
 CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
 CC the invention are used to treat or prevent infections by a pathogenic
 CC organism such as A. fumigatus, to treat a non-infectious disease in a

CC subject (e.g. cancer), to prevent or contain contamination of an object
 CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
 CC biofilm comprising A. fumigatus. The polynucleotides are useful for
 CC expressing recombinant protein for characterisation, screening or
 CC therapeutic use, as markers for host tissues in which the pathogenic
 CC organisms invade or reside, for comparing with the DNA sequence of A.
 CC fumigatus to identify duplicated genes or paralogues having the same or
 CC similar biochemical activity and/or function, for comparisons with DNA
 CC sequences of other related or distant pathogenic organisms to identify
 CC potential orthologous essential or virulence genes, for selecting and
 CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune
 CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC elicit immune response, as a reagent in assays designed to quantitatively
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organism invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This sequence represents a protein of one of the essential genes
 CC of Aspergillus fumigatus of the invention.
 XX
 SQ Sequence 546 AA;
 Query Match 0.8%; Score 8; DB 24; Length 546;
 Best Local Similarity 100.0%; Pred. No. 80;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 540 LMGHEYIE 547
 Db 326 LMGHEYIE 333
 |||||
 RESULT 11
 AAY30210
 ID AAY30210 standard; Protein; 687 AA.
 XX
 AC AAY30210;
 XX
 XX 01-NOV-1999 (first entry)
 DT
 XX Amino acid sequence of a saccharide conversion enzyme.
 XX
 XX Saccharide conversion enzyme; trehalose synthase activity;
 KW Paedomonas sp. Fl.
 KW
 XX Pseudomonas sp.
 OS
 XX JP11225769-A.
 PN
 XX 24-AUG-1999.
 PD
 XX 19-FEB-1998; 98JP-0054433.
 PF
 XX 19-FEB-1998; 98JP-0054433.
 PR
 XX (SHIZ-) SHIZUOKA KEN.
 PA
 XX WPI; 1999-521084/44.
 DR N-PSDB; AA210213, AA210214.
 DR
 XX New DNA encoding saccharide conversion enzyme - can be used to
 PT prepare poly:peptide commercially
 PT
 XX Claim 4; Page 12-14; 22pp; Japanese.
 PS
 XX The present sequence represents a saccharide conversion enzyme. The
 CC protein has trehalose synthase activity, and is isolated from
 CC Paedomonas sp. Fl. The specification describes methods for producing
 CC the enzyme commercially.
 XX
 XX Sequence 687 AA;
 SQ

Query Match 0.8%; Score 8; DB 20; Length 687;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 859 SVTPPTIG 866
 DB 233 SVTPPTIG 240

RESULT 12
 ABJ26326
 ID ABJ26326 standard; Protein; 689 AA.
 AC ABJ26326;
 DT 16-APR-2003 (first entry)
 DE Aspergillus fumigatus essential gene protein #984.
 XX Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
 KW cancer; contamination; biofilm; antibody; immune response.
 XX Aspergillus fumigatus.
 OS WO200286090-A2.
 XX 31-OCT-2002.
 XX 23-APR-2002; 2002WO-US11142.
 XX 23-APR-2001; 2001US-285697P.
 PR 27-APR-2001; 2001US-287066P.
 PR 05-JUN-2001; 2001US-295890P.
 PR 09-JUL-2001; 2001US-303899P.
 PR 31-AUG-2001; 2001US-316362P.
 XX (ELIT-) ELITRA PHARM INC.
 XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
 XX WPI; 2003-093124/08.
 XX New purified or isolated nucleic acids of essential genes of
 PT Aspergillus fumigatus, useful for treating or preventing infections by
 PT A. fumigatus, or for treating a non-infectious disease in a subject
 PT e.g. cancer
 XX Disclosure; Page -, 175pp; English.

The invention relates to novel purified or isolated nucleic acids of essential genes of Aspergillus fumigatus. The isolated nucleic acids of the invention are used to treat or prevent infections by a pathogenic organism such as A. fumigatus, to treat a non-infectious disease in a subject (e.g. cancer), to prevent or contain contamination of an object by A. fumigatus, or to prevent or inhibit formation on a surface of a biofilm comprising A. fumigatus. The polynucleotides are useful for expressing recombinant protein for characterization, screening or therapeutic use, as markers for host tissues in which the pathogenic organisms invade or reside, for comparing with the DNA sequence of A. fumigatus to identify duplicated genes or paralogues having the same or similar biochemical activity and/or function, for comparing with DNA sequences of other related or distant pathogenic organisms to identify potential orthologous essential or virulence genes, for selecting and making oligomers for attachment to a nucleic acid array for examination of expression patterns, for raising anti-protein antibodies, as an antigen to raise anti-DNA antibodies or to elicit another immune response, and for identifying polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction. The polypeptides may be used to raise antibodies or to elicit immune response, as a reagent in assays designed to quantitatively determine levels of the protein in biological fluids, as a marker for host tissues in which pathogenic organism invade or reside, and to

CC isolate correlative receptors or ligands in the case or virulence factors. This sequence represents a protein of one of the essential genes of Aspergillus fumigatus of the invention.

XX
 SQ Sequence 689 AA;
 Query Match 0.8%; Score 8; DB 24; Length 689;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 LMGHEYIE 547
 DB 433 LMGHEYIE 440

RESULT 13
 ABB90939
 ID ABB90939 standard; Protein; 1043 AA.
 AC ABB90939;
 DT 31-MAY-2002 (first entry)
 XX Herbicidally active polypeptide SEQ ID NO 150.
 DE Herbicidal; plant; agriculture; herbicide.
 XX Arabidopsis thaliana.
 XX WO200210210-A2.
 XX 07-FEB-2002.
 XX 28-AUG-2001; 2001WO-EP09892.
 XX 28-AUG-2001; 2001WO-EP09892.
 XX (FARB) BAYER AG.
 XX Tietjen K, Weidler M;
 XX WPI; 2002-269010/31.
 XX Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms -
 XX Claim 5; SEQ ID NO 150; 261pp + Sequence Listing; English.

The invention relates to identifying target proteins (ABB90790-ABB94016) for herbicidally active compounds, comprising aligning and comparing nucleic acid or amino acid sequences from plant with nucleic acid or amino acid sequences from non-plant organisms using suitable search parameters, where plant sequences having an E-value greater by a factor of 3 than the E-value of most similar non-plant sequences are selected. The polypeptides or nucleic acids encoding them are useful for identifying modulators. The identified modulators are useful as herbicides.

XX
 SQ Sequence 1043 AA;
 Query Match 0.8%; Score 8; DB 23; Length 1043;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 821 IDKKTGKQ 828
 DB 980 IDKKTGKQ 987

RESULT 14
 AAW04505

ID AAW04505 standard; Protein; 1992 AA.
 AC AAW04505;
 XX
 DT 25-JAN-1997 (first entry)
 XX
 DE Moraxella 200 kDa outer membrane protein.
 XX
 KW Outer membrane protein; OMP; immunogen; vaccine; otitis media;
 KW diagnosis.
 XX
 KW Mycobacterium catarrhalis strain 4223.
 XX
 OS WO9634960-A1.
 PN
 XX 07-NOV-1996.
 PD
 XX 29-APR-1996; 96WO-CA00264.
 XX
 XX 26-MAR-1996; 96US-0621944.
 PR
 PR 01-MAY-1995; 95US-0431718.
 PR
 XX 07-JUN-1995; 95US-0478370.
 XX
 PA (CONN-) CONNAUGHT LAB LTD.
 XX
 XX Chong P, Harkness RE, Klein MH, Loosmore SM, Sasaki K;
 PI
 XX WPI; 1996-506162/50.
 DR
 DR N-PSDB; AAT38740.
 XX
 XX Moraxella outer membrane protein - useful as immunogen in protective
 PT vaccine and for diagnosis
 XX
 PS Claim 14; Fig 6; 109pp; English.
 XX
 CC An approx. 200 kDa outer membrane protein (AAW04505) can be
 CC isolated from Moraxella catarrhalis otitis media strain 4223
 CC by electroelution, or expressed from a gene (see also AAT38740)
 CC obtd. from a strain 4223 genomic library. Natural or recombinant
 CC outer membrane protein is useful as an immunogen to protect
 CC against infection by Moraxella, esp. M. catarrhalis. It can
 CC also be used to detect antibodies, esp. for differential diagnosis
 CC between bacteria that cause similar symptoms, and also useful as
 CC a carrier for other antigens and used to raise antitumour
 CC antibodies for conjugation to therapeutic agents.
 XX
 SQ Sequence 1992 AA;
 Query Match 0.8%; Score 8; DB 17; Length 1992;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 836 VDADGNKV 843
 DB 1337 VDADGNKV 1344
 RESULT 15
 AAB69133
 ID AAB69133 standard; Protein; 1992 AA.
 XX
 AC AAB69133;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa protein SEQ ID NO:3.
 XX
 KW Moraxella catarrhalis strain 4223; major outer membrane protein;
 KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;
 KW otitis media; Detection.
 XX
 OS Moraxella catarrhalis.
 XX

PN WO200107619-A1.
 XX
 PD 01-FEB-2001.
 XX
 PF 26-JUL-2000; 2000WO-CA00870.
 XX
 XX 27-JUL-1999; 99US-0361619.
 PR
 XX (CONN-) CONNAUGHT LAB LTD.
 PA
 XX Loosmore SM, Sasaki K, Yang Y, Klein MH;
 PI
 XX WPI; 2001-159722/16.
 DR
 DR N-PSDB; AAF59100, AAF59101.
 XX
 PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
 XX useful in protective vaccines and for diagnosis -
 XX
 PS Example 3; Fig 2A-W; 247pp; English.
 XX
 CC The present invention describes an isolated and purified nucleic acid (I)
 CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
 CC The 200 kDa outer membrane protein (II) has antibacterial activity and
 CC can be used in vaccines. (II), and its truncated versions, are used as
 CC immunogenic compositions and vaccines to protect against M. catarrhalis
 CC infections, particularly otitis media in humans. (II) is also used as
 CC antigen in immunoassays for detecting specific antibodies (Ab), and to
 CC generate Ab. (I) are used for recombinant production of (II) and its
 CC fragments are used as probes for identifying/cloning 200 kDa protein
 CC genes from other strains, and for diagnostic detection of M. catarrhalis.
 CC (I) makes possible production of large amount of recombinant immunogens.
 CC Expression of truncated versions of (II) reduces toxicity of the protein
 CC towards the Escherichia coli host. The present sequence represents the
 CC M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa protein, which is
 CC used in the exemplification of the present invention.
 XX
 SQ Sequence 1992 AA;
 Query Match 0.8%; Score 8; DB 22; Length 1992;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 836 VDADGNKV 843
 DB 1337 VDADGNKV 1344

Search completed: January 7, 2004, 19:10:33
 Job time : 49 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2004, 19:08:48 ; Search time 27 Seconds
(without alignments)
3622.354 Million cell updates/sec

Title: US-09-581-286A-424
Perfect score: 1017
Sequence: 1 MKRWTLFLCLLTSGWAMA.....VGKNQYPNSKQYVAGIQLSF 1017

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 9616862 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 76:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	0.8	183	2 T47251	complex I protein
2	8	0.8	183	2 A36621	NADH2 dehydrogenase
3	8	0.8	217	2 A12186	hypothetical prote
4	8	0.8	321	1 A28557	chloride peroxidase
5	8	0.8	340	2 C69154	homoserine dehydro
6	8	0.8	342	2 AH1071	conserved hypothet
7	8	0.8	343	2 A98295	ribosomal RNA smal
8	8	0.8	343	2 C86136	probable enzyme yj
9	8	0.8	343	2 S58595	RNA (guanine-N2-)
10	8	0.8	487	2 B93059	hypothetical prote
11	8	0.8	487	2 A97928	type I site-specif
12	8	0.8	501	1 JN0539	head protein gp5 -
13	8	0.8	501	2 D90767	probable head-tail
14	8	0.8	501	2 D90970	probable head-tail
15	8	0.8	501	2 D85717	hypothetical prote
16	8	0.8	501	2 C85743	probable head-tail
17	8	0.8	607	2 F64227	DNA primase (dnaB)
18	8	0.8	683	1 S69780	outer membrane pro
19	8	0.8	763	2 T21006	hypothetical prote
20	8	0.8	820	2 A86510	leucyl tRNA synthet
21	8	0.8	820	2 C72113	leucine-tRNA ligase
22	8	0.8	1006	2 E96683	hypothetical prote
23	8	0.8	1038	2 JC6027	115K outer membran
24	8	0.8	1096	2 H86237	protein F14N23.29
25	7	0.7	30	2 PH0249	T-cell receptor Vb
26	7	0.7	41	2 S19600	Alc9Y protein - mo
27	7	0.7	67	2 B98067	degenerate transpo
28	7	0.7	79	2 T42000	hypothetical prote
29	7	0.7	104	2 T36424	probable ABC-type

30	7	0.7	105	2 H97566	hypothetical prote
31	7	0.7	105	2 AF2787	hypothetical prote
32	7	0.7	118	2 B71059	hypothetical prote
33	7	0.7	141	2 S18397	hemoglobin alpha c
34	7	0.7	149	2 F70599	hypothetical prote
35	7	0.7	153	2 I39827	hypothetical prote
36	7	0.7	154	2 A82812	conserved hypothet
37	7	0.7	155	2 AE1602	conserved hypothet
38	7	0.7	155	2 A11239	B. subtilis YlxS p
39	7	0.7	156	2 B96523	P1A17.13 [importe
40	7	0.7	156	2 A32868	prolactin receptor
41	7	0.7	171	2 E87512	dihydrofolate redu
42	7	0.7	177	2 H83323	hypothetical prote
43	7	0.7	179	2 G89854	conserved hypothet
44	7	0.7	180	2 S39489	mannose-binding le
45	7	0.7	182	2 AD2024	hypothetical prote

ALIGNMENTS

RESULT 1

T47251 complex I protein 22K [imported] - Neurospora crassa

C:Species: Neurospora crassa

C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 21-Jul-2000

C:Accession: T47251

R:Videira, A.; Tropischueg, M.; Wachter, E.; Schneider, H.; Werner, S.

J. Biol. Chem. 265, 13060-13065, 1990

A:Title: Molecular cloning of subunits of complex I from Neurospora crassa. Primary stru

A:Reference number: A36621; MUID:90330647; PMID:2142943

A:Accession: T47251

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-183 <VID>

A:Cross-references: EMBL:M55323; NID:gl68780; PIDN:AAA35571.1; PID:gl68781

Query Match 0.8%; Score 8; DB 2; Length 183;

Best Local Similarity 100.0%; Pred. No. 7.2;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 LGASSAPL 179

DB 29 LGASSAPL 36

RESULT 2

A36621

NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) 22K chain precursor - Neurospora crassa

N:Alternate names: complex I protein 22K chain

C:Species: Neurospora crassa

C>Date: 28-Mar-1991 #sequence_revision 30-Jan-1993 #text_change 03-Jun-2002

C:Accession: A36621

R:Videira, A.; Tropischueg, M.; Wachter, E.; Schneider, H.; Werner, S.

J. Biol. Chem. 265, 13060-13065, 1990

A:Title: Molecular cloning of subunits of complex I from Neurospora crassa. Primary stru

A:Reference number: A36621; MUID:90330647; PMID:2142943

A:Accession: A36621

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-183 <VID>

A:Cross-references: GB:J05559

A>Note: the authors translated the codon CGG for residue 46 as Pro and CTT for residue 7

C:Keywords: mitochondrion; NAD; oxidoreductase

Query Match

Best Local Similarity 100.0%; Pred. No. 7.2;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 LGASSAPL 179

DB 29 LGASSAPL 36

RESULT 3

AI2186
hypothetical protein all3048 [imported] - Nostoc sp. (strain PCC 7120)
A;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AI2186
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AI2186
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-217 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAE74747.1; PID:g17132142; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all3048

Query Match 0.8%; Score 8; DB 2; Length 217;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 386 GANLSGAI 393

Db 202 GANLSGAI 209

RESULT 4

A28557
Chloride peroxidase (EC 1.11.1.10) precursor - fungus (Leptoxiphium fumago)
N;Alternate names: chloroperoxidase
C;Species: Leptoxiphium fumago
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
A;Accession: A28557; A25548; A25936; A60824; A31956
R;Nuell, M.J.; Pang, G.H.; Axley, M.J.; Kenigsberg, P.; Hager, L.P.
J. Bacteriol. 170, 1007-1011, 1988
A;Title: Isolation and nucleotide sequence of the chloroperoxidase gene from Caldariomyces fumago
A;Reference number: A28557; MUID:88115133; PMID:2828306
A;Accession: A28557
A;Molecule type: DNA
A;Residues: 1-321 <NUE>
A;Cross-references: GB:M19025; GB:X04486; NID:g167240; PIDN:AAA33026.1; PID:g167241
R;Fang, G.H.; Kenigsberg, P.; Axley, M.J.; Nuell, M.; Hager, L.P.
Nucleic Acids Res. 14, 8061-8071, 1986
A;Title: Cloning and sequencing of chloroperoxidase cDNA.
A;Reference number: A25548; MUID:87040773; PMID:3774552
A;Accession: A25548
A;Molecule type: mRNA
A;Residues: 1-321 <FAN>
A;Cross-references: GB:X04486; NID:g2555; PIDN:CAA28172.1; PID:g642215
R;Axley, M.J.; Kenigsberg, P.; Hager, L.P.
J. Biol. Chem. 261, 15058-15061, 1986
A;Title: Fructose induces and glucose represses chloroperoxidase mRNA levels.
A;Reference number: A25936; MUID:87033742; PMID:3771564
A;Accession: A25936
A;Molecule type: protein
A;Residues: 1-176 <AXL>
R;Kenigsberg, P.; Fang, G.H.; Hager, L.P.
Arch. Biochem. Biophys. 254, 409-415, 1987
A;Title: Post-translational modifications of chloroperoxidase from Caldariomyces fumago.
A;Reference number: A60824; MUID:87212023; PMID:2883934
A;Accession: A60824
A;Molecule type: protein
A;Residues: 22-24, 'A', 26-32, 'DD', 35-40, 'A', 42-44, 'T' <KEN>
A;Note: 25-Ser, 33-Asn, and 34-Asn were also found
A;Note: two Asn residues, including 34-Asn, and one Gln residue were deamidated during isolation
A;Note: 22-Glu appeared to cyclize to pyrrolidone carboxylic acid during isolation
R;Blanke, S.R.; Hager, L.P.

J. Biol. Chem. 263, 18739-18743, 1988

A;Title: Identification of the fifth axial heme ligand of chloroperoxidase.
A;Reference number: A31956; MUID:89066662; PMID:3198598
A;Accession: A31956

A;Molecule type: protein

A;Residues: 39-54;91-98, 'D', 100-113, 'V', 115-124 <BLA>

C;Superfamily: chloroperoxidase

C;Keywords: chromoprotein; extracellular protein; glycoprotein; heme; iron; metalloproteins

F;1-21/Domain: signal sequence #status predicted <SIG>

F;22-321/Product: chloride peroxidase #status experimental <MAT>

F;33/Binding site: carboxylate (Asn) (covalent) #status experimental

F;50/Binding site: heme iron (Cys) (axial ligand) #status experimental

F;100-108/Disulfide bonds: #status experimental

Query Match 0.8%; Score 8; DB 1; Length 321;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 QTSLDVVA 193

Db 157 QTSLDVVA 164

RESULT 5

C69154
homoserine dehydrogenase homolog - Methanobacterium thermoautotrophicum (strain Delta H)
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 11-Jun-1999
C;Accession: C69154
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Olu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functions and genome organization
A;Reference number: A69000; MUID:98037514; PMID:9371463
A;Accession: C69154
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-340 <MTH>

A;Cross-references: GB:AE000826; GB:AE000666; NID:g2621478; PIDN:AA84923.1; PID:g2621481

A;Experimental source: strain Delta H

C;Genetics:

A;Gene: MTH417

A;Start codon: GTG

C;Superfamily: homoserine dehydrogenase; homoserine dehydrogenase homology

F;1-264/Domain: homoserine dehydrogenase homology <HSD>

Query Match 0.8%; Score 8; DB 2; Length 340;

Best Local Similarity 100.0%; Pred. No. 13;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 260 MTGDELLD 267

Db 79 MTGDELLD 86

RESULT 6

AH1071
conserved hypothetical protein STY4906 [imported] - Salmonella enterica subsp. enterica
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AH1071
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AH1071
A;Status: preliminary
A;Molecule type: DNA

A;Residues: 1-342 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD03391.1; PID:g16505660; GSPDB:GN00176
C;Genetics:
A;Gene: STY4906

Query Match 0.8%; Score 8; DB 2; Length 342;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 DGMQTSLD 190

|||||
Db 273 DGMQTSLD 280

RESULT 7
A98295
ribosomal RNA small subunit methyltransferase [imported] - Escherichia coli (strain O157
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: A98295
R;Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A;Reference number: A98295; MUID:21156231; PMID:11258796
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-343 <HAY>
A;Cross-references: GB:BA000007; PIDN:BA838752.1; PID:g13364807; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs5329

Query Match 0.8%; Score 8; DB 2; Length 343;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 DGMQTSLD 190

|||||
Db 273 DGMQTSLD 280

RESULT 8
C86136
probable enzyme YjJt [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: C86136
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: C86136
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-343 <STO>
A;Cross-references: GB:AE005174; NID:g12519394; PIDN:AAG59551.1; GSPDB:GN00145; UWGP:Z59
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: YjJt

Query Match 0.8%; Score 8; DB 2; Length 343;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 DGMQTSLD 190

|||||
Db 273 DGMQTSLD 280

RESULT 9

S56595

rRNA (guanine-N2-) methyltransferase (EC 2.1.1.52) - Escherichia coli (strain K-12)
N;Alternate names: hypothetical protein f343b
C;Species: Escherichia coli
C;Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 01-Mar-2002
C;Accession: S56595; B65252
R;Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res. 23, 2105-2119, 1995
A;Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92
A;Reference number: S56314; MUID:95334362; PMID:7610040
A;Accession: S56595

A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA

A;Residues: 1-343 <BUR>
A;Cross-references: EMBL:U14003; NID:g1263172; PIDN:AAA97267.1; PID:g537211
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: B65252

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-343 <BLAT>

A;Cross-references: GB:AE000507; GB:U00096; NID:g2367380; PIDN:AAC77324.1; PID:g1790830,
A;Experimental source: strain K-12, substrain MG1655

C;Genetics:

A;Gene: yjJt

C;Keywords: methyltransferase

Query Match 0.8%; Score 8; DB 2; Length 343;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 DGMQTSLD 190

|||||
Db 273 DGMQTSLD 280

RESULT 10

B95059
hypothetical protein SP0509 [imported] - Streptococcus pneumoniae (strain TIGR4)
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 16-Aug-2002
C;Accession: B95059
R;Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001

A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A;Reference number: A95000; MUID:21357209; PMID:11463916

A;Accession: B95059

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-487 <KUR>

A;Cross-references: GB:AE005672; PIDN:AAK74667.1; PID:g14971982; GSPDB:GN00164; TIGR:SP4

A;Experimental source: strain TIGR4

C;Genetics:

A;Gene: SP0509

C;Superfamily: type I site-specific deoxyribonuclease chain hsdM

Query Match 0.8%; Score 8; DB 2; Length 487;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 261 TGDLELDF 268

|||||
Db 77 TGDLELDF 84

RESULT 11

A97928
 type I site-specific deoxyribonuclease (EC 3.1.21.3) chain hsdM [imported] - Streptococcus
 C;Species: Streptococcus pneumoniae
 C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 16-Aug-2002
 C;Accession: A97928
 R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; H
 e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
 y, P.; Sun, P.M.; Winkler, M.E.
 J. Bacteriol. 183, 5709-5717, 2001
 A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
 A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
 A;Reference number: A97872; MUID:21429245; PMID:11544234
 A;Accession: A97928
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-487 <KUR>
 A;Cross-references: GB:AE007317; PIDN:AAK99253.1; PID:g15458017; GSPDB:GN00174
 C;Genetics:
 A;Gene: hsdM
 C;Superfamily: type I site-specific deoxyribonuclease chain hsdM

Query Match 0.8%; Score 8; DB 2; Length 487;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 261 TGDLELDF 268
 |||||
 Db 77 TGDLELDF 84

RESULT 12
 JN0539
 head protein gp5 - phase 21
 N;Contains: head protein gp6
 C;Species: phase 21
 C;Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 23-Jul-1999
 C;Accession: JN0539; JN0540
 R;Smith, M.P.; Reiss, M.
 Gene 126, 1-7, 1993
 A;Title: Sequence analysis of the phase 21 genes for prohead assembly and head completio
 A;Reference number: JN0537; MUID:93231520; PMID:8472949
 A;Accession: JN0539
 A;Molecule type: DNA
 A;Residues: 1-501 <SMI>
 A;Cross-references: GB:M81255; NID:g215454; PIDN:AAA32343.1; PID:g215459
 C;Genetics:
 A;Gene: 5
 A;Start codon: GTG
 C;Superfamily: phage lambda minor capsid protein C
 C;Keywords: capsid protein; head protein
 F;302-501/Product: head protein gp6 #status predicted <GP6>

Query Match 0.8%; Score 8; DB 1; Length 501;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 480 AQAGVDIT 487
 |||||
 Db 199 AQAGVDIT 206

RESULT 13
 D90767
 probable head-tail preconnector protein [imported] - Escherichia coli (strain O157:H7, s
 C;Species: Escherichia coli
 C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
 C;Accession: D90767
 R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gend
 A;Reference number: A99629; MUID:21156231; PMID:11258796
 A;Accession: D90767

A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-501 <HAY>
 A;Cross-references: GB:BA000007; PIDN:BA834531.1; PID:g13360568; GSPDB:GN00154
 A;Experimental source: strain O157:H7, substrain RIMD 0509952
 C;Genetics:
 A;Gene: ECe1108
 C;Superfamily: phage lambda minor capsid protein C

Query Match 0.8%; Score 8; DB 2; Length 501;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 480 AQAGVDIT 487
 |||||
 Db 199 AQAGVDIT 206

RESULT 14
 D90970
 probable head-tail preconnector protein [imported] - Escherichia coli (strain O157:H7, s
 C;Species: Escherichia coli
 C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
 C;Accession: D90970
 R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom
 A;Reference number: A99629; MUID:21156231; PMID:11258796
 A;Accession: D90970
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-501 <HAY>
 A;Cross-references: GB:BA000007; PIDN:BA836155.1; PID:g13362200; GSPDB:GN00154
 A;Experimental source: strain O157:H7, substrain RIMD 0509952
 C;Genetics:
 A;Gene: ECe2732
 C;Superfamily: phage lambda minor capsid protein C

Query Match 0.8%; Score 8; DB 2; Length 501;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 480 AQAGVDIT 487
 |||||
 Db 199 AQAGVDIT 206

RESULT 15
 D85717
 hypothetical protein Z2134 [imported] - Escherichia coli (strain O157:H7, substrain EDL9;
 C;Species: Escherichia coli
 C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C;Accession: D85717
 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousia, K.; Apodaca,
 Nature 409, 529-533, 2001
 A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A;Reference number: A85480; MUID:21074935; PMID:11206551
 A;Accession: D85717
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-501 <STO>
 A;Cross-references: GB:AE005174; NID:g12515088; PIDN:AAG56200.1; GSPDB:GN00145; UWGP:Z21;
 A;Experimental source: strain O157:H7, substrain EDL933
 C;Genetics:
 A;Gene: Z2134
 C;Superfamily: phage lambda minor capsid protein C

Query Match 0.8%; Score 8; DB 2; Length 501;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 480 AQAGVDIT 487
| | | | |
Db 199 AQAGVDIT 206

Search completed: January 7, 2004, 19:12:43
Job time : 30 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2004, 19:05:03 ; Search time 18 Seconds
(without alignments)
2657.009 Million cell updates/sec

Title: US-09-581-286A-424
Perfect score: 1017
Sequence: 1 MKRMTLFLCLLTSGWAMA.....VGKNQPNKQYVAGIQLSF 1017

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	8	0.8	183	1 NUPM_NEUCR
2	8	0.8	342	1 RSMC_ECOLI
3	8	0.8	373	1 PRXC_CALFU
4	8	0.8	501	1 VG05_BPP21
5	8	0.8	607	1 PRIM_MYCB
6	8	0.8	820	1 SYL_CHLPN
7	7	0.7	141	1 HBA_ECHTE
8	7	0.7	151	1 RNB_HSV2H
9	7	0.7	153	1 YCT3_BACFI
10	7	0.7	169	1 TCTP_SCHJA
11	7	0.7	195	1 YV23_CABEL
12	7	0.7	212	1 GIDB_XANCF
13	7	0.7	239	1 CALD_MELGA
14	7	0.7	249	1 ULA3_HCMVA
15	7	0.7	261	1 YX00_MYCLE
16	7	0.7	274	1 YTXM_BACSU
17	7	0.7	307	1 PYRB_HELPJ
18	7	0.7	307	1 PYRB_HELPY
19	7	0.7	314	1 OGW1_HUMAN
20	7	0.7	359	1 HAH4_HALME
21	7	0.7	364	1 TRU2_THETN
22	7	0.7	372	1 TOLA_HAEIN
23	7	0.7	381	1 PBPA_NOCIA
24	7	0.7	389	1 PSD6_HUMAN
25	7	0.7	389	1 PSD6_MOUSE
26	7	0.7	400	1 FD12_MORAP
27	7	0.7	400	1 TGT_MORIS
28	7	0.7	400	1 TGT_CAEEL
29	7	0.7	402	1 PURA_BRUAB
30	7	0.7	403	1 CID_DROYA
31	7	0.7	404	1 ILV5_SCHPO
32	7	0.7	412	1 YWFA_BACSU
33	7	0.7	415	1 YBDG_ECOLI
				P21976 neurospora
				P39406 escherichia
				P46963 caldaxiomyc
				P36273 bacterioph
				P47492 mycoplasma
				O92930 chlamydia p
				P24291 echinops te
				P89479 herpes simp
				Q04445 bacillus fi
				P91800 schistosoma
				P50437 caenorhabdi
				O8930 xanthomonas
				P13505 melesagris g
				P16734 human cytom
				Q49741 mycobacteri
				P23974 bacillus su
				O92m81 helicobacte
				Q25716 helicobacte
				Q96rr8 homo sapien
				Q48236 halobacteri
				Q8x7f0 thermomanae
				P44678 haemophilus
				Q06317 nocardia la
				Q15008 homo sapien
				Q99j14 mus musculu
				Q9Y8h5 mortierella
				P59668 mortierella
				Q23623 caenorhabdi
				P52004 brucella ab
				O77027 drosophila
				P78827 schizosacch
				P39637 bacillus su
				P39455 escherichia

34	7	0.7	421	1 POK_EMENI
35	7	0.7	423	1 IDH_BACSU
36	7	0.7	429	1 PURA_BRUME
37	7	0.7	432	1 PURA_RHILO
38	7	0.7	438	1 DNAB_TREPA
39	7	0.7	447	1 PURA_SYNEL
40	7	0.7	449	1 YN60_YEAST
41	7	0.7	462	1 NORM_BURVI
42	7	0.7	471	1 SAHH_CHLTE
43	7	0.7	473	1 CYP1_CYNCA
44	7	0.7	478	1 IF3Y_YEAST
45	7	0.7	478	1 VTNC_HUMAN

ALIGNMENTS

RESULT 1
ID NUPM_NEUCR STANDARD; PRT; 183 AA.
AC P21976;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE NADH-ubiquinone oxidoreductase 20.8 kDa subunit (EC 1.6.5.3)
DE (EC 1.6.99.3).
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-45.
RX MEDLINE=90330647; PubMed=2142943;
RA Videira A., Tropschug M., Wachter E., Schneider H., Werner S.;
RT "Molecular cloning of subunits of complex I from Neurospora crassa.
Primary structure and in vitro expression of a 22-kDa polypeptide.";
J. Biol. Chem. 265:13060-13065(1990).
CC -!- FUNCTION: TRANSFER OF ELECTRONS FROM NADH TO THE RESPIRATORY
CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED
TO BE UBIQUINONE.
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -!- CATALYTIC ACTIVITY: NADH + acceptor = NAD(+) + reduced acceptor.
CC -!- COFACTOR: Binds 1 iron-sulfur cluster (Potential).
CC -!- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 30 DIFFERENT SUBUNITS.
CC THIS IS A COMPONENT OF THE HYDROPHOBIC FRACTION.
CC -!- SUBCELLULAR LOCATION: Matrix and cytoplasmic side of the
mitochondrial inner membrane.
CC -!- SIMILARITY: BELONGS TO THE COMPLEX I 19 kDa SUBUNIT FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch)

EMBL; M55323; AAA33571.1; -;
DR PIR; T47251; T47251.
SQ Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
KW SEQUENCE 183 AA; 20911 MW; A2574693F41093D4 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 183;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 172 LGASSAPL 179
|||||||
Db 29 LGASSAPL 36

RESULT 2
RSMC_ECOLI

ID RSMC ECOLI STANDARD; PRT; 342 AA.
 AC P39406;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DE Ribosomal RNA small subunit methyltransferase C (EC 2.1.1.52) (rRNA
 DE (guanine-N(2)-methyltransferase) (16S rRNA m2G1207
 DE methyltransferase).
 GN RSMC OR B4371.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=95334362; PubMed=7610040;
 RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
 RA Blattner F.R.;
 RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
 RT region from 92.8 through 100 minutes.";
 RL Nucleic Acids Res. 23:2105-2119(1995).
 RN [2]
 RP SEQUENCE OF 1-19, AND CHARACTERIZATION.
 RX MEDLINE=99091632; PubMed=9873033;
 RA Tcherne J.S., Nurse K., Popienick P., Ofengand J.;
 RT "Purification, cloning, and characterization of the 16S RNA m2G1207
 RT methyltransferase from Escherichia coli.";
 RL J. Biol. Chem. 274:924-929(1999).
 CC -1- FUNCTION: SPECIFICALLY METHYLATES THE GUANOSINE IN POSITION 1207
 CC OF 16S RNA IN THE 30S PARTICLE.
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + rRNA = S-adenosyl-L-
 CC homocysteine + rRNA containing N(2)-methylguanine.
 CC -1- COFACTOR: REQUIRES MAGNESIUM FOR ACTIVITY.
 CC -1- SIMILARITY: BELONGS TO THE METHYLTRANSFERASE SUPERFAMILY. RSMC
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; U14003; AAA97267.1; -;
 DR EMBL; AB00507; AAC77324.1; -;
 DR PIR; S56595; S56595.
 DR EcoGene; EG12596; ramC.
 DR InterPro; IPR002296; NL2N6_mtfraae.
 DR InterPro; IPR002052; N6_Mcrae.
 DR Pfam; PF05175; MTS; 2.
 DR PRINTS; PR00507; N12N6MTFRASE.
 DR PROSITE; PS00092; N6_MTFASE; UNKNOWN 1.
 KW rRNA processing; Transferase; Methyltransferase; Magnesium;
 KW complete proteome.
 FT INIT MET 0
 SQ SEQUENCE 342 AA; 37493 MW; C7A318155700302D CRC64;
 Query Match 0.8%; Score 8; DB 1; Length 342;
 Best Local Similarity 100.0%; Pred. No. 9.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 183 DGMQTSLD 190
 Db 272 DGMQTSLD 279
 RESULT 3
 PRXC CALFU STANDARD; PRT; 373 AA.
 ID PRXC CALFU
 AC P04963; Q92216; Q9HPP2;

DT 13-AUG-1987 (Rel. 05, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chloroperoxidase precursor (EC 1.11.1.10) (Chloride peroxidase) (CPO).
 GN CPO.
 OS Caldariomyces fumago (Leptoxiphium fumago).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
 OC Capnoidiales; Capnodiaceae; mitosporic Capnodiaceae; Leptoxiphium.
 OX NCBI_TaxID=5474;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=ATCC 16373;
 RX MEDLINE=87040773; PubMed=3774552;
 RA Fang G.-H., Kenigsberg P., Axley M.J., Nuell M., Hager L.P.;
 RT "Cloning and sequencing of chloroperoxidase cDNA.";
 RL Nucleic Acids Res. 14:8061-8071(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88115133; PubMed=2828306;
 RA Nuell M.J., Fang G.-H., Axley M.J., Kenigsberg P., Hager L.P.;
 RT "Isolation and nucleotide sequence of the chloroperoxidase gene from
 RT Caldariomyces fumago.";
 RL J. Bacteriol. 170:1007-1011(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21264537; PubMed=11278701;
 RA Conesa A., an de Velde P., van Rantwijk F., Sheldon R.A.,
 RA van den Hondel C.A.M.J.J., Funt P.J.;
 RT "Expression of the Caldariomyces fumago Chloroperoxidase in
 RT Aspergillus niger and characterization of the recombinant enzyme.";
 RL J. Biol. Chem. 276:17635-17640(2001).
 RN [4]
 RP SEQUENCE OF 1-175 FROM N.A.
 RX MEDLINE=87033742; PubMed=3771564;
 RA Axley M.J., Kenigsberg P., Hager L.P.;
 RT "Fructose induces and glucose represses chloroperoxidase mRNA
 RT levels.";
 RL J. Biol. Chem. 261:15058-15061(1986).
 RN [5]
 RP REVISIONS TO C-TERMINUS.
 RA Hager L.P.;
 RL Unpublished observations (FEB-1996).
 RN [6]
 RP DETERMINATION OF HEME LIGAND.
 RX MEDLINE=89066662; PubMed=3198598;
 RA Blanke S.R., Hager L.P.;
 RT "Identification of the fifth axial heme ligand of chloroperoxidase.";
 RL J. Biol. Chem. 263:18739-18743(1988).
 RN [7]
 RP MUTAGENESIS OF CYS-50.
 RX MEDLINE=20006248; PubMed=10535936;
 RA Yi X., Mroczko M., Manoj K.M., Wang X., Hager L.P.;
 RT "Replacement of the proximal heme thiolate ligand in chloroperoxidase
 RT with a histidine residue.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:12412-12417(1999).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
 RC STRAIN=ATCC 16373;
 RX MEDLINE=96363674; PubMed=8747463;
 RA Sundaramoorthy M., Turner J., Poulos T.L.;
 RT "The crystal structure of chloroperoxidase: a heme peroxidase-
 RT cytochrome P450 functional hybrid.";
 RL Structure 3:1387-1377(1995).
 CC -1- FUNCTION: Catalyzes peroxidative halogenations involved in the
 CC biosynthesis of clardariomycin (2,2-dichloro-1,3-cyclo-
 CC pentenedione). The enzyme also has potent catalase activity and in
 CC the absence of halide ion, acts as a peroxidase similar to plant
 CC peroxidases.
 CC -1- CATALYTIC ACTIVITY: 2 RH + 2 chloride + H(2)O(2) = 2 RCl + 2
 CC H(2)O.
 CC -1- COFACTOR: Binds 1 protoheme IX, 1 iron(III) ion and 1 manganese
 CC ion per subunit.
 CC -1- PTM: N- AND O-GLYCOSYLATED.

Head	Gaps	0;
;		
d		
IBLY		
IT IS		
laboration		
ststation -		
ns on its		
n no way		
commercial		
/announce/		

```
FT REPEAT 418 455 TRI3.
SQ SEQUENCE 501 AA; 52440 MW; 1C7C4A450A09FCFC CRC64;

Query Match 0.8%; Score 8; DB 1; Length 501;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 480 AQAGVDIT 487
|||||
Db 199 AQAGVDIT 206

RESULT 5
PRIM_MYCGE STANDARD; PRT; 607 AA.
AC P47432;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA Primase (EC 2.7.7.-).
GN DNAG OR DNAE OR MG250.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.M., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-P., Dougherty B.A., Brott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RA "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403 (1995).
RN [2]
RC SEQUENCE OF 78-216 AND 511-607 FROM N.A.
RP STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RL J. Bacteriol. 175:7918-7930 (1993).
CC -1- FUNCTION: DNA PRIMASE IS THE POLYMERASE THAT SYNTHESIZES SMALL
CC RNA PRIMERS FOR THE OKAZAKI FRAGMENTS ON BOTH TEMPLATE STRANDS AT
CC REPLICATION FORKS DURING CHROMOSOMAL DNA SYNTHESIS.
CC -1- COFACTOR: BINDS ONE ZINC ION PER MOLECULE (BY SIMILARITY).
CC -1- SUBUNIT: Monomer (By similarity).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U39704; AAC71470.1; -
CC EMBL; U02146; AAD12426.1; -
CC EMBL; U01771; AAD10589.1; -
CC PIR; F64227; F64227.
CC TIGR; MG250; -
CC InterPro; IPR006295; DNA_primase.
CC InterPro; IPR006171; Toprim_dom.
CC InterPro; IPR006647; Toprim_primase.
CC InterPro; IPR006154; Toprim_sub.
CC InterPro; IPR002694; Znf_CHC2.
CC Pfam; PF01751; Toprim_1.
CC Pfam; PF01807; zf-CHC2; 1.
CC ProDom; PD002276; Toprim_primase; 1.
CC ProDom; PD002988; Znf_CHC2; 1.
CC SMART; SM00493; TOPRIM; 1.

DR SMART; SM00400; Znf_CHCC; 1.
DR TIGRFAMs; TIGR01391; dnaG; 1.
KW Transferrase; DNA replication; DNA-directed RNA polymerase; Primosome;
KW Zinc-finger; Zinc; Metal-binding; Complete proteome.
FT ZN_FING 39 63 CHC2-TYPE (BY SIMILARITY).
SQ SEQUENCE 607 AA; 71061 MW; A84730CF8A6BDEA CRC64;

Query Match 0.8%; Score 8; DB 1; Length 607;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 948 RLKNLKLT 955
|||||
Db 599 RLKNLKLT 606

RESULT 6
SYL_CHLPN STANDARD; PRT; 820 AA.
AC Q9Z930; Q9JQ86;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS OR CPN0153 OR CP0618.
OS Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX NCBI_TaxID=83558;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S., trachomatis.";
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389 (1999).
RN [2]
RC SEQUENCE FROM N.A.
RP STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA White T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F., Bass S.,
RA Read O., Hickey E.K., Peterson J., Utterback T., Berry K., Nakazawa T.;
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406 (2000).
RN [3]
RC SEQUENCE FROM N.A.
RP STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314 (2000).
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB001602; AAD18306.1; -
CC EMBL; AB002219; AAF38433.1; -
CC EMBL; AF002545; BAA98363.1; -
```


DR PIR; A86510; A86510.
 DR PIR; C72113; C72113.
 DR TIGR; CP0618; -.
 DR HAMAP; MF_00049; -, 1.
 DR InterPro; IPR002302; Leu-TRNA-synt1a.
 DR InterPro; IPR002300; tRNA-synt_1a.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR Pfam; PF00133; tRNA-synt_1; 1.
 DR PRINTS; PRO0985; TRNASYNTHLEU.
 DR TIGRFAMs; TIGR00396; leuS bact; 1.
 DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Complete proteome.
 FT SITE 40 51 "HIGH" REGION.
 FT SITE 601 605 "RMSKS" REGION.
 FT BINDING 604 604 ATP (BY SIMILARITY).
 SQ SEQUENCE 820 AA; 93965 MW; 520369FC098F1926 CRC64;
 Query Match 0.8%; Score 8; DB 1; Length 820;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 105 VVLGYGTG 112
 DB 325 VVLGYGTG 332
 RESULT 7
 HBA_ECHTE STANDARD; PRT; 141 AA.
 ID P24291;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hemoglobin alpha chain.
 GN HBA.
 OS Echinops telfairi (Lesser hedgehog tenrec).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Insectivora; Tenrecidae; Tenrecinae; Echinops.
 OX NCBI_TaxID=9371;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=92172283; PubMed=1793518;
 RA Piccinini M., Kleinschmidt T., Gorr T., Weber R.E., Kuenzle H.,
 Braunitzer G.;
 RT "Primary structure and oxygen-binding properties of the hemoglobin
 from the lesser hedgehog tenrec (Echinops telfairi, Zalambdodonta).
 Evidence for phylogenetic isolation.";
 RL Biol. Chem. Hoppe-Seyler 372:975-989(1991).
 CC -!- FUNCTION: Involved in oxygen transport from the lung to the
 various peripheral tissues.
 CC -!- SUBUNIT: Heterotetramer of two alpha chains and two beta chains.
 CC TWO EXTERNAL CYSTEINE RESIDUES AT BETA-16 AND BETA-52 CAUSE
 REVERSIBLE POLYMERIZATION TO OCTAMERS AND MOST LIKELY
 IRREVERSIBLE FORMATION OF HIGHER POLYMERS.
 CC -!- TISSUE SPECIFICITY: Red blood cells.
 CC -!- SIMILARITY: Belongs to the globin family.
 DR PIR; S18397; S18397.
 DR HSSP; P01958; 2MHE.
 DR InterPro; IPR002338; Alpha haem.
 DR InterPro; IPR000971; Globin.
 DR Pfam; PF00042; Globin; 1.
 DR PRINTS; PRO0612; ALPHAHAE.
 DR PROSITE; PS01033; GLOBIN; 1.
 KW Heme; Oxygen transport; Transport; Erythrocyte.
 FT METAL 58 58 IRON (HEME DISTAL LIGAND).
 FT METAL 87 87 IRON (HEME PROXIMAL LIGAND).
 SQ SEQUENCE 141 AA; 15027 MW; 379F8241EC1E9D29 CRC64;
 Query Match 0.7%; Score 7; DB 1; Length 141;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 994 AGGNVGK 1000
 DB 17 AGGNVGK 23
 RESULT 8
 RNB_HSV2H STANDARD; PRT; 151 AA.
 ID P89479;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Potential RNA-binding protein.
 GN US11.
 OS Herpes simplex virus (type 2 / strain HG52).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Simplexvirus.
 OX NCBI_TaxID=10315;
 RN [1]
 RP SEQUENCE FROM N.A.
 RL Dolan A.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: BINDS DNA AND RNA (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar (by similarity).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL; Z86099; CAB06719.1; -
 DR DNA-binding; RNA-binding; Repeat; Nuclear protein.
 KW DOMAIN 90 146 11 X 6 AA TANDEM REPEATS.
 FT REPEAT 90 95 1.
 FT REPEAT 96 101 2.
 FT REPEAT 102 104 3.
 FT REPEAT 105 110 4.
 FT REPEAT 111 116 5.
 FT REPEAT 117 122 6.
 FT REPEAT 123 128 7.
 FT REPEAT 129 130 8.
 FT REPEAT 131 134 9.
 FT REPEAT 135 140 10.
 FT REPEAT 141 146 11.
 SQ SEQUENCE 151 AA; 16297 MW; FAB751F23C3DB6AE CRC64;
 Query Match 0.7%; Score 7; DB 1; Length 151;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 758 DLSLKGK 764
 DB 20 DLSLKGK 26
 RESULT 9
 YCT3_BACFI STANDARD; PRT; 153 AA.
 ID YCT3_BACFI
 AC Q04445;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 17.8 kDa protein in ctap 3' region (ORF3).
 OS Bacillus firmus.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1399;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OF4;
 RX MEDLINE=93107080; PubMed=7678007;

RA Quirk P.G., Hicks D.B., Krulwich T.A.;
 RT "Cloning of the cta operon from alkaliphilic *Bacillus firmus* OF4 and
 RT characterization of the pH-regulated cytochrome caa3 oxidase it
 RT encodes";
 RL J. Biol. Chem. 268:678-685(1993).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M94110; AAA22370.1; -
 CC Hypothetical protein.
 KW Hypothetical protein.
 SQ SEQUENCE 153 AA; 17748 MW; DAB8C2BE453B67CF CRC64;
 . Query Match 0.7%; Score 7; DB 1; Length 153;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 264 ELLDFQV 270
 Db 91 ELLDFQV 97
 RESULT 10
 TCCTP SCHJA STANDARD; PRT; 169 AA.
 ID TCCTP SCHJA
 AC P91800;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Translationally controlled tumor protein homolog (TCCTP).
 GN TCCTP.
 OS Schistosoma japonicum (Blood fluke).
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
 OC Schistosomatidae; Schistosomatidae; Schistosoma.
 OX NCBI_TaxID=6182;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Chinese;
 RA Fan J., Drew A., Brindley P.J.;
 RA "Characterization of a cDNA encoding a homolog of translationally
 RT controlled tumor protein (TCCTP) in *S. japonicum*";
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE TCCTP FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U85483; AAB42079.1; -
 DR HSPSP; Q10344; Ith6Q.
 DR InterPro; IPR001983; TCCTP.
 DR Pfam; PF00838; TCCTP; 1.
 DR ProDom; PD004329; TCCTP; 1.
 DR PROSITE; PS01002; TCCTP_1; 1.
 DR PROSITE; PS01003; TCCTP_2; 1.
 SQ SEQUENCE 169 AA; 19376 MW; 3C7355765C9DB13B CRC64;
 . Query Match 0.7%; Score 7; DB 1; Length 169;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 375 FESRINE 381
 |||||

Db 111 FESRINE 117
 RESULT 11
 YY23 CAEEL STANDARD; PRT; 195 AA.
 ID YY23 CAEEL
 AC P50437;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Hypothetical 20.2 kDa protein C28C12.3 in chromosome IV.
 GN C28C12.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Miller N.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U40797; AAB37546.1; -
 DR PIR; T15679;
 DR WormPep; C28C12.3; CE04110.
 DR InterPro; IPR003677; Onchocerca_Ag.
 DR Pfam; PF02520; DUF148; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 195 AA; 20183 MW; 59C2D7523E45ACD7 CRC64;
 . Query Match 0.7%; Score 7; DB 1; Length 195;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 764 TIYQNKD 770
 Db 122 TIYQNKD 128
 |||||
 RESULT 12
 GIDB XANCP STANDARD; PRT; 212 AA.
 ID GIDB XANCP
 AC Q8P3N0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Methyltransferase gldb (SC 2.1.-) (Glucose inhibited division
 DE protein B).
 GN GIDB OR XCC4041.
 OS Xanthomonas campestris (pv. campestris).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=340;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33913 / NCPPB 528;
 RX MEDLINE=22022145; PubMed=12024217;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.P.,
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
 RA Camarotte G., Cannavan F., Cardozo J., Chamberg F., Ciapina L.P.,
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,

RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Okura C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities";
RL Nature 417:459-463(2002).
CC -1- FUNCTION: Probable S-adenosyl-L-methionine dependent
CC methyltransferase specific for a sterol and/or lipid substrate (By
CC similarity). BELONGS TO THE GIDB FAMILY.
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE012526; AAA43262.1; -;
DR HAMAP; MF_00074; -; 1.
DR InterPro; IPR003682; GIDB.
DR Pfam; PF02527; GIDB; 1.
DR ProDom; PD004441; GIDB; 1.
DR TIGRFAMs; TIGR00138; GIDB; 1.
KW Transferase; Methyltransferase; Complete proteome.
SQ SEQUENCE 212 AA; 22682 MW; 338940B291F60E31 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 212;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 878 SLDADPA 884
Db 22 SLDADPA 28
|||||

RESULT 13
CALD_MELGA
ID CALD MELGA STANDARD; PRT; 239 AA.
AC P13505;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Caldesmon, smooth muscle (CDM) (Fragment).
GN CALDI OR CAD.
OS Meleagris gallopavo (Common turkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Meleagrididae; Meleagris.
OX NCBI_TaxID=9103;
RN [1]
RP SEQUENCE.
RC TISSUE=Gizzard;
RX MEDLINE=89273603; PubMed=2730648;
RA Leszyk J., Mornet D., Audemard E., Collins J.H.;
RT "Caldesmon structure and function: sequence analysis of a 35
RT kilodalton actin- and calmodulin-binding fragment from the C-terminus
RT of the turkey gizzard protein.";
RL Biochem. Biophys. Res. Commun. 160:1371-1378(1989).
RN [2]
RP SEQUENCE OF 1-96.
RC TISSUE=Gizzard;
RX MEDLINE=89228015; PubMed=2653315;
RA Leszyk J., Mornet D., Audemard E., Collins J.H.;
RT "Amino acid sequence of a 15 kilodalton actin-binding fragment of
RT turkey gizzard caldesmon: similarity with dystrophin, tropomyosin and
RT the tropomyosin-binding region of troponin T.";
RL Biochem. Biophys. Res. Commun. 160:210-216(1989).
CC -1- FUNCTION: CONTROL OF ACTOMYOSIN INTERACTIONS IN SMOOTH MUSCLE

CC AND NONMUSCLE CELLS (COULD ACT AS A BRIDGE BETWEEN MYOSIN AND
CC ACTIN FILAMENTS). INHIBITS THE ACTIN-ACTIVATED ATPASE OF MYOSIN
CC THIS INHIBITION IS ATTENUATED BY CALCIUM-CALMODULIN AND IS
CC POTENTIATED BY TROPOMYOSIN. INTERACTS WITH ACTIN, MYOSIN,
CC 2 MOLECULES OF TROPOMYOSIN AND WITH CALMODULIN
CC -1- SUBCELLULAR LOCATION: ON THIN FILAMENTS IN SMOOTH MUSCLE AND ON
CC STRESS FIBERS IN FIBROBLASTS (NONMUSCLE).
CC PIR; S16925; S16925.
DR InterPro; IPR006017; Caldesmon.
DR Pfam; PF02029; Caldesmon; 1.
DR PRINTS; PRO1076; CALDESMON.
KW Muscle protein; Actin-binding; Calmodulin-binding.
FT NON_TER 1
FT 239 239
SQ SEQUENCE 239 AA; 26561 MW; D395BBB8B465B8895 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 239;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 377 SRINEWL 383
Db 205 SRINEWL 211
|||||

RESULT 14
ULGA3_HCMVA
ID ULGA3_HCMVA STANDARD; PRT; 249 AA.
AC P16734;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Protein UL103.
GN UL103.
OS Human cytomegalovirus (strain AD169).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90269039; PubMed=2161319;
RA Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,
RA Horsnell T., Hutchison C.A. III, Kouzarides T., Martignetti J.A.,
RA Freddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;
RT "Analysis of the protein-coding content of the sequence of human
RT cytomegalovirus strain AD169";
RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL7,
CC EBV-1 55, VZV 53, EBV BBRF2, HCMV UL103 AND HSV-1 42.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X17403; CAA35339.1; -;
DR PIR; S09868; S09868.
DR InterPro; IPR002600; Herpes UL7.
DR Pfam; PF01677; Herpes UL7; 1.
SQ SEQUENCE 249 AA; 28636 MW; A8D9F8F89F02FE9D CRC64;

Query Match 0.7%; Score 7; DB 1; Length 249;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 647 LSDRLK 653
Db 83 LSDRLK 89
|||||

```
RESULT 15
YY00 MYCLE
ID YY00 MYCLE STANDARD; PRT; 261 AA.
AC Q49741;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein ML0393.
GN ML0393 OR B1620_F3_119.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R., Robison K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Horneby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 403:1007-1013(2001).
CC -1- SIMILARITY: STRONG, TO M.TUBERCULOSIS RV3400 AND SOME, TO
CC M.TUBERCULOSIS MTCY39.11C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U00015; AAC43246.1; -.
DR EMBL; AL583918; CAC29901.1; -.
DR PIR; S72837; S72837.
DR Leproma; ML0393; -.
DR InterPro; IPR005834; Hydrolase.
DR Pfam; PF00702; Hydrolase; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 261 AA; 28545 MW; E8F8A7357B53CFAB CRC64;
Query Match 0.7%; Score 7; DB 1; Length 261;
Best Local Similarity 100.0%; Pred.No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 39 IGANVVV 45
Db 248 IGANVVV 254
|||||
|||||
```

Search completed: January 7, 2004, 19:11:04
Job time : 20 secs

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	815	80.1	1017	2	Q9ZA60	Q9ZA60 porphyromon
2	39	3.8	59	2	Q9XAU9	Q9XAU9 porphyromon
3	9	0.9	508	2	Q9S4M1	Q9S4M1 escherichia
4	9	0.9	521	2	Q9XC53	Q9XC53 escherichia
5	9	0.9	523	2	Q9XC54	Q9XC54 escherichia
6	8	0.8	217	16	Q8YSN6	Q8YSN6 anabaena sp
7	8	0.8	275	5	Q9U3J0	Q9U3J0 caenorhabdi
8	8	0.8	340	17	Q2G517	Q2G517 methanobact
9	8	0.8	342	16	Q8ZJW6	Q8ZJW6 salmonella
10	8	0.8	342	16	Q8Z0V2	Q8Z0V2 salmonella
11	8	0.8	343	16	Q8X510	Q8X510 salmonella
12	8	0.8	343	16	Q8FA64	Q8FA64 escherichia
13	8	0.8	413	5	Q8MLM1	Q8MLM1 drosophila
14	8	0.8	435	5	Q8T0D5	Q8T0D5 drosophila
15	8	0.8	487	16	Q97S79	Q97S79 streptococc
16	8	0.8	487	16	Q8DX11	Q8DX11 streptococc

```
QY 121 SVAKVSSEKLAEPVANIIMDALQOVAGMOWMTTSGDPTAVASVEIHGTGSLGASSAPLY 180
DB 121 SVAKVSSEKLAEPVANIIMDALQOVAGMOWMTTSGDPTAVASVEIHGTGSLGASSAPLY 180
QY 181 IVDMQTSLSLDVATMNPNDPESMSVLKASATSIIYGARAANGVVFIQTKGKMSERGRIT 240
DB 181 IVDMQTSLSLDVATMNPNDPESMSVLKASATSIIYGARAANGVVFIQTKGKMSERGRIT 240
QY 241 FNASYGIGSOILNTKPLDMMTGDDELLDPQVAGFWGNNQTVQVKOMILAGAEDLYGYND 300
DB 241 FNASYGIGSOILNTKPLDMMTGDDELLDPQVAGFWGNNQTVQVKOMILAGAEDLYGYND 300
QY 301 SLKDEYKTLFPVDFNDHDAWLKALPKTAPTQSGDISFSGSQSTSYASIGYFDQSGMA 360
DB 301 SLKDEYKTLFPVDFNDHDAWLKALPKTAPTQSGDISFSGSQSTSYASIGYFDQSGMA 360
QY 361 REPANFKRYSGRLNFESRINEMLKVAGNLGAIANRRSADYFGKYNGSGTGGVLTMPRY 420
DB 361 REPANFKRYSGRLNFESRINEMLKVAGNLGAIANRRSADYFGKYNGSGTGGVLTMPRY 420
QY 421 YNPDPVNGDLADVYTMGATRPSMTBYPFAKMRPFSSSHOANVNGFAQITPIKGLTLKA 480
DB 421 YNPDPVNGDLADVYTMGATRPSMTBYPFAKMRPFSSSHOANVNGFAQITPIKGLTLKA 480
QY 481 QAGVDITNTRTSSKRMPPNPYDSTPLGERRERAYRDVSKSFTNTAEYKFSIDEKHDLTAL 540
DB 481 QAGVDITNTRTSSKRMPPNPYDSTPLGERRERAYRDVSKSFTNTAEYKFSIDEKHDLTAL 540
QY 541 MGHEYIEYEGDVI GASSKGPFESDKMLLSQKTCNSLSLPEHRVAEYAYLSFSPSRFNYGF 600
DB 541 MGHEYIEYEGDVI GASSKGPFESDKMLLSQKTCNSLSLPEHRVAEYAYLSFSPSRFNYGF 600
QY 601 DKWYIDFSVRNDQSSRFGSNRNSAWFYSGGMPDIYNKFIQESNWLSDRLKMSYGTGG 660
DB 601 DKWYIDFSVRNDQSSRFGSNRNSAWFYSGGMPDIYNKFIQESNWLSDRLKMSYGTGG 660
QY 661 NSEIGNYHQAALVTNNYTDAMGLSISTAGNPDLSWEKOSQFNFLAAGAFNNRLSAEV 720
DB 661 NSEIGNYHQAALVTNNYTDAMGLSISTAGNPDLSWEKOSQFNFLAAGAFNNRLSAEV 720
QY 721 DFYVTRTNDMLIDVPMPIYISGFFSQYQNVGSMKNTGVDLSLKGTYIQNKDNNVYASAFN 780
DB 721 DFYVTRTNDMLIDVPMPIYISGFFSQYQNVGSMKNTGVDLSLKGTYIQNKDNNVYASAFN 780
QY 781 YNROEITKLPFGLNKYMLPNTGTIWEIGYNSFYMABYAGIDKKTGKQLWYVPGQVDADG 840
DB 781 YNROEITKLPFGLNKYMLPNTGTIWEIGYNSFYMABYAGIDKKTGKQLWYVPGQVDADG 840
QY 841 NKVTTQYSADLETRIDKSVTPPTTGGPSLGASWKGSLDADFAIYVKGWMINNDRYFTE 900
DB 841 NKVTTQYSADLETRIDKSVTPPTTGGPSLGASWKGSLDADFAIYVKGWMINNDRYFTE 900
QY 901 NAGGLMQLNKDMLNNAWTDENKEDTVPKLGQSPQFDTHLENASFLRLKNLKLTYVLPN 960
DB 901 NAGGLMQLNKDMLNNAWTDENKEDTVPKLGQSPQFDTHLENASFLRLKNLKLTYVLPN 960
QY 961 SLFAGQNVIGGARVYLMARNLLTVTKYKGFDPDEAGNGVGNQYQNSQYVAGIQLSF 1017
DB 961 SLFAGQNVIGGARVYLMARNLLTVTKYKGFDPDEAGNGVGNQYQNSQYVAGIQLSF 1017
RESULT 2
Q9XAU9 PRELIMINARY; PRT; 59 AA.
AC Q9XAU9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 13, Last annotation update)
DE RgaA protein (Fragment).
GN RGA.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
```

```
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RX MEDLINE=20316027; PubMed=10858216;
RA Bonase W.A., Marsh P.D., Percival R.S., Aduse-Opoku J., Hanley S.A.,
RA Devine D.A., Curtis M.A.;
RT "Identification of rgaAB as a temperature-regulated operon of
RT Porphyromonas gingivalis W50 using differential display of randomly
RT primed RNA.";
RL Infect. Immun. 68:4012-4017(2000).
DR EMBL; AJ242673; CAB46019.1; -.
FT NON_TER 1 1
FT NON_TER 59 59
SQ SEQUENCE 59 AA; 6486 MW; D3527513DABB33AC CRC64;
Query Match 3.8%; Score 39; DB 2; Length 59;
Best Local Similarity 100.0%; Pred. No. 7.4e-32;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 429 DLADVYTMGATRPSMTBYPFAKMRPFSSSHOANVNGF 467
DB 1 DLADVYTMGATRPSMTBYPFAKMRPFSSSHOANVNGF 39
RESULT 3
Q9S4M1 PRELIMINARY; PRT; 508 AA.
ID Q9S4M1
AC Q9S4M1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Flagellin (Fragment).
GN FLIC.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Bacterichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=E851/71;
RX MEDLINE=99084952; PubMed=9864325;
RA Reid S.D., Selander R.K., Whittam T.S.;
RT "Sequence, diversity of flagellin (fliC) alleles in pathogenic
RT Escherichia coli.";
RL J. Bacteriol. 181:153-160(1999).
DR EMBL; AF128955; AAD28526.2; -.
DR InterPro; IPR001492; FlagellinN.
DR InterPro; IPR001029; Flagellin_C.
DR Pfam; PF00700; Flagellin_C; 1.
DR PRINTS; PD00207; Flagellin_N; 1.
DR PRINTS; PD000316; Flagellin_C; 2.
FT NON_TER 1 1
FT NON_TER 508 508
SQ SEQUENCE 508 AA; 52098 MW; 8CB41D64572A8FCB CRC64;
Query Match 0.9%; Score 9; DB 2; Length 508;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 838 ADGNKVTTTS 846
DB 209 ADGNKVTTTS 217
RESULT 4
Q9XCS3 PRELIMINARY; PRT; 521 AA.
ID Q9XCS3
AC Q9XCS3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
```

DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Flagellin (fragment).
GN FLIC.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DEC 2a;
RX MEDLINE=99084952; PubMed=9864325;
RA Reid S.D., Selander R.K., Whittam T.S.;
RT "Sequence diversity of flagellin (flc) alleles in pathogenic
RT Escherichia coli.";
RL J. Bacteriol. 181:153-160(1999).
DR EMBL; AF128949; AAD28520.2; -.
DR InterPro; IPR001492; FlagellinN.
DR InterPro; IPR001029; FlagellinC.
DR Pfam; PF00700; Flagellin_C; 1.
DR Pfam; PF00669; Flagellin_C; 1.
DR PRINTS; PR00207; FLAGELLIN.
DR ProDom; PD000316; Flagellin_C; 2.
FT NON_TER 1
FT NON_TER 521
SQ SEQUENCE 521 AA; 53266 MW; C9EF9BD5944AC848 CRC64;

Query Match 0.9%; Score 9; DB 2; Length 521;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 838 ADGNKVITS 846
|||||
DB 208 ADGNKVITS 216

RESULT 5
QXACS4
ID Q9XCS4 PRELIMINARY; PRT; 523 AA.
AC Q9XCS4;
RX STRAIN=DEC 1a;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Flagellin (fragment).
GN FLIC.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DEC 1a;
RX MEDLINE=99084952; PubMed=9864325;
RA Reid S.D., Selander R.K., Whittam T.S.;
RT "Sequence diversity of flagellin (flc) alleles in pathogenic
RT Escherichia coli.";
RL J. Bacteriol. 181:153-160(1999).
DR EMBL; AF128948; AAD28519.2; -.
DR InterPro; IPR001492; FlagellinN.
DR InterPro; IPR001029; FlagellinC.
DR Pfam; PF00700; Flagellin_C; 1.
DR Pfam; PF00669; Flagellin_C; 1.
DR PRINTS; PR00207; FLAGELLIN.
DR ProDom; PD000316; Flagellin_C; 2.
FT NON_TER 1
FT NON_TER 523
SQ SEQUENCE 523 AA; 53621 MW; 0FCA8E31D82C8D73 CRC64;

Query Match 0.9%; Score 9; DB 2; Length 523;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 838 ADGNKVITS 846
|||||

Db 213 ADGNKVITS 221

RESULT 6
Q8YSN6
ID Q8YSN6 PRELIMINARY; PRT; 217 AA.
AC Q8YSN6;
RX STRAIN=DEC 2a;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Hypothetical protein A113048.
GN A113048.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yaeuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003591; BAB74747.1; -.
DR InterPro; IPR001646; Speptide_repeat.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00226; DnaJ; 1.
DR Pfam; PF00805; Pentapeptide; 2.
DR SMART; SM00271; DnaJ; 1.
DR PROSITE; PS00076; DnaJ_2; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 217 AA; 24211 MW; AD7EAE4D85A7C1AB CRC64;

Query Match 0.8%; Score 8; DB 16; Length 217;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 386 GANLSGAI 393
|||||
DB 202 GANLSGAI 209

RESULT 7
Q9U3J0
ID Q9U3J0 PRELIMINARY; PRT; 275 AA.
AC Q9U3J0;
RX STRAIN=DEC 1a;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE F16A11.3 protein.
GN F16A11.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Baynes C.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
RL EMBL; Z81505; CAB60288.1; -.
DR WormPep; F16A11.3; CE23664.
SQ SEQUENCE 275 AA; 31604 MW; 6D80D2EDF15B6617 CRC64;

```

Query Match          0.8%; Score 8; DB 5; Length 275;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 KLSTVSGS 121
DB 76 KLSTVSGS 83

RESULT 8
O26517 PRELIMINARY; PRT; 340 AA.
AC O26517;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Homoserine dehydrogenase homolog.
GN MTH417.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwan N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA Daniels C.J., Mao J.-i., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AE000826; AAB84923.1; -
DR InterPro; IPR001342; Homoserine dh.
DR Pfam; PF00742; Homoserine_3.
DR Pfam; PF03447; NAD_binding_3; 1.
DR PROSITE; PS01042; HOMOSER_DHGENASE; 1.
KW Complete proteome.
SQ SEQUENCE 340 AA; 35952 MW; F6FAED73691A29E CRC64;

Query Match          0.8%; Score 8; DB 17; Length 340;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 260 MTGDELLD 267
DB 79 MTGDELLD 86

RESULT 9
O82JW6 PRELIMINARY; PRT; 342 AA.
AC O82JW6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 16S rRNA m2G 1207 methylase.
GN RSMC OR STM4556.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,

```

```

RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan G., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
DR EMBL; AE008914; AAL23371.1; -.
DR InterPro; IPR002052; N6_Mtase.
DR InterPro; IPR000051; SAM_bind.
DR PROSITE; PS00092; N6_MTASE; 1.
KW Complete proteome.
SQ SEQUENCE 342 AA; 37615 MW; F034BFD4CBEBE00F CRC64;

Query Match          0.8%; Score 8; DB 16; Length 342;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 DGMQTSILD 190
DB 273 DGMQTSILD 280

RESULT 10
O82OV2 PRELIMINARY; PRT; 342 AA.
AC O82OV2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein STY4906.
GN STY4906.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
DR EMBL; AL627284; CAD003391.1; -.
DR InterPro; IPR002052; N6_Mtase.
DR InterPro; IPR000051; SAM_bind.
DR PROSITE; PS00092; N6_MTASE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 342 AA; 37625 MW; E2A480C9AE17F124 CRC64;

Query Match          0.8%; Score 8; DB 16; Length 342;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 DGMQTSILD 190
DB 273 DGMQTSILD 280

RESULT 11
O8X510 PRELIMINARY; PRT; 343 AA.
AC O8X510;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

```


DE Putative enzyme (Ribosomal RNA small subunit methyltransferase).
GN YJUT OR Z5972 OR EC5329.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12";
RL DNA Res. 8:11-22(2001).
DR EMBL; AE005668; AGS5551.1; -;
DR EMBL; AP002569; BAB38752.1; -;
DR InterPro; IPR002052; N6_Mtase.
DR InterPro; IPR000051; SAM_bind.
DR PROSITE; PS00092; N6_MTASE; 1.
KW Complete proteome.
SQ SEQUENCE 343 AA; 37601 MW; E145C3F1121104A0 CRC64;

Query Match 0.8%; Score 8; DB 16; Length 343;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 DGMQTSLD 190
DB 273 DGMQTSLD 280
|||||

RESULT 12
Q8FA64 PRELIMINARY; PRT; 343 AA.
AC Q8FA64;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Ribosomal RNA small subunit methyltransferase C (EC 2.1.1.52).
GN YJUT OR C5450.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016772; AAN83870.1; -;
KW Methyltransferase; Transferase; Complete proteome.
SQ SEQUENCE 343 AA; 37639 MW; 5DF09B83DC3CF857 CRC64;

Query Match 0.8%; Score 8; DB 16; Length 343;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 DGMQTSLD 190
DB 273 DGMQTSLD 280
|||||

RESULT 13
Q8MLM1 PRELIMINARY; PRT; 413 AA.
AC Q8MLM1;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE CG30438-PA.
GN Drosophila melanogaster (Fruit fly).
OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Kravitz S., Kulp D., Lai Z.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon K.Y., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,

RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskin R.A., Hostin D., Howland T.J.,
RA Ilegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacieb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svitek R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RP Miara S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kanink J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RN SEQUENCE FROM N.A.
RP Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RN SEQUENCE FROM N.A.
RA Flybase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AEO03786; AAM68364.1; -;
DR Flybase; FBgn050438; CG30438.
DR InterPro; IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
SQ SEQUENCE 413 AA; 47439 MW; 95BB865ED4029C72 CRC64;

Query Match 0.8%; Score 8; DB 5; Length 413;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 686 SISTAGNP 693
Db 51 SISTAGNP 58
|||||

RESULT 14
Q8TOD5 PRELIMINARY; PRT; 435 AA.
ID Q8TOD5
AC Q8TOD5
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE LD09936p.
GN BCDNA:LD09936.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacieb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY069392; AAL39537.1; -;
DR Flybase; FBgn0047220; BCDNA:LD09936.
DR InterPro; IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.

SQ SEQUENCE 435 AA; 49909 MW; EAC08C44A699F1B7 CRC64;

Query Match 0.8%; Score 8; DB 5; Length 435;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 686 SISTAGNP 693
Db 73 SISTAGNP 80
|||||

RESULT 15
Q97S79 PRELIMINARY; PRT; 487 AA.
ID Q97S79
AC Q97S79
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Type I restriction-modification system, M subunit.
GN SP0509.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapfel E., Khouri H., Wolf A.M., Ueberback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae";
RL Science 293:498-506(2001).
DR EMBL; AEO07362; AAK74667.1; -;
DR TIGR; SP0509; -;
DR InterPro; IPR003665; Methylase_M.
DR InterPro; IPR002296; N12N6_mtfase.
DR InterPro; IPR003356; N6_DNA_Mtase.
DR InterPro; IPR000051; SAM_bind.
DR Pfam; PF02506; Methylase_M; 1.
DR Pfam; PF02384; N6_Mtase; 1.
DR PRINTS; PR00507; N12NGMTFRASE.
KW Complete proteome.
SQ SEQUENCE 487 AA; 56539 MW; 782137154BF6B249 CRC64;

Query Match 0.8%; Score 8; DB 16; Length 487;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 261 TGDELLDF 268
Db 77 TGDELLDF 84
|||||

Search completed: January 7, 2004, 19:12:03
Job time : 48 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2004, 19:09:44 ; Search time 22 Seconds
(without alignments)
1955.916 Million cell updates/sec

Title: US-09-581-286A-424
Perfect score: 1017
Sequence: 1 MKRMTLFLCLLTSGNAMA.....VGNQYNSKQYVAGIQLSF 1017

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0 *No word size was specified.*

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents_AA*
1: /cgm2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgm2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgm2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgm2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgm2_6/ptodata/1/1aa/PTCUS_COMB.pep:*
6: /cgm2_6/ptodata/1/1aa/backfiles.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	0.8	295	4	US-09-328-352-6656
2	8	0.8	338	4	US-09-107-532A-5819
3	8	0.8	507	4	US-09-198-452A-174
4	8	0.8	1833	4	US-08-621-944A-4
5	8	0.8	1833	4	US-08-945-567D-4
6	8	0.8	1992	4	US-08-621-944A-3
7	8	0.8	1992	4	US-08-945-567D-3
8	8	0.8	2123	3	US-08-968-685A-10
9	8	0.8	2314	4	US-09-268-347-49
10	8	0.7	10	6	5252328-2
11	7	0.7	41	3	US-08-737-629-6
12	7	0.7	77	3	US-08-686-968C-2
13	7	0.7	88	1	US-08-285-440-27
14	7	0.7	88	1	US-08-285-440-27
15	7	0.7	90	1	US-08-630-349-25
16	7	0.7	90	1	US-08-285-440-25
17	7	0.7	94	1	US-08-630-349-25
18	7	0.7	94	1	US-08-285-440-22
19	7	0.7	111	4	US-08-630-349-22
20	7	0.7	116	1	US-08-936-165A-437
21	7	0.7	116	1	US-08-285-440-1
22	7	0.7	116	1	US-08-630-349-1
23	7	0.7	118	1	US-08-285-440-2
24	7	0.7	118	1	US-08-630-349-2
25	7	0.7	122	1	US-08-285-440-3
26	7	0.7	122	1	US-08-630-349-3
27	7	0.7	207	4	US-09-252-991A-31339
			241	6	5213972-7
					Patent No. 5213972

28	7	0.7	253	4	US-09-071-035-12	Sequence 12, Appl
29	7	0.7	255	4	US-09-614-474-2	Sequence 2, Appl
30	7	0.7	272	4	US-09-071-035-10	Sequence 10, Appl
31	7	0.7	291	4	US-09-107-532A-7140	Sequence 7140, Ap
32	7	0.7	296	4	US-09-328-352-6102	Sequence 6102, Ap
33	7	0.7	312	1	US-08-285-440-4	Sequence 4, Appl
34	7	0.7	312	1	US-08-630-349-4	Sequence 4, Appl
35	7	0.7	356	4	US-09-107-532A-5906	Sequence 5906, Ap
36	7	0.7	397	4	US-09-252-991A-28422	Sequence 28422, A
37	7	0.7	399	2	US-08-834-655-4	Sequence 4, Appl
38	7	0.7	399	3	US-08-834-033A-4	Sequence 4, Appl
39	7	0.7	399	3	US-09-363-574-4	Sequence 4, Appl
40	7	0.7	399	4	US-09-363-526-4	Sequence 4, Appl
41	7	0.7	399	4	US-09-330-235-20	Sequence 20, Appl
42	7	0.7	420	4	US-09-328-352-4729	Sequence 4729, Ap
43	7	0.7	428	4	US-09-134-001C-3801	Sequence 3801, Ap
44	7	0.7	447	4	US-09-252-991A-32755	Sequence 32755, A
45	7	0.7	476	1	US-08-216-276A-33	Sequence 33, Appl

ALIGNMENTS

RESULT 1
US-09-328-352-6656
; Sequence 6656, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6656
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6656

Query Match 0.8%; Score 8; DB 4; Length 295;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	125	VSSKLAIE	132
Db	280	VSSKLAIE	287

RESULT 2
US-09-107-532A-5819
; Sequence 5819, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A

```
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5819:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 338 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...338
; SEQUENCE DESCRIPTION: SEQ ID NO: 5819:
US-09-107-532A-5819

Query Match 0.8%; Score 8; DB 4; Length 338;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 VSSEKLA 132
DB 268 VSSEKLA 275

RESULT 3
US-09-198-452A-174
; Sequence 174, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffois, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 174
; LENGTH: 507
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: SITE
; LOCATION: 1...507
; OTHER INFORMATION: Xaa=unknown or other
US-09-198-452A-174

Query Match 0.8%; Score 8; DB 4; Length 507;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 VVLGYGTG 112
DB 325 VVLGYGTG 332

RESULT 4
US-08-621-944A-4
; Sequence 4, Application US/08621944A
; Patent No. 6440425
```

```
; GENERAL INFORMATION:
; APPLICANT: SASAKI, Ken
; APPLICANT: HARKNESS, Robin E.
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE PROTEIN OF MORAXELLA
; TITLE OF INVENTION: MEMBRANE PROTEIN OF MORAXELLA
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/621,944A
; FILING DATE: 26-MAR-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/478,370
; FILING DATE: 07-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-587
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1833 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-621-944A-4

Query Match 0.8%; Score 8; DB 4; Length 1833;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 836 VDADGNKV 843
DB 1178 VDADGNKV 1185

RESULT 5
US-08-945-567D-4
; Sequence 4, Application US/08945567D
; Patent No. 6448386
; GENERAL INFORMATION:
; APPLICANT: SASAKI, Ken
; APPLICANT: HARKNESS, Robin E.
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: CHONG, Pele
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-745 MIS
; CURRENT APPLICATION NUMBER: US/08/945,567D
; CURRENT FILING DATE: 1996-04-29
; PRIOR APPLICATION NUMBER: 08/431,718
; PRIOR FILING DATE: 1995-05-01
; PRIOR APPLICATION NUMBER: 08/478,370
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 08/621,944
; PRIOR FILING DATE: 1996-03-26
```

;; PRIOR APPLICATION NUMBER: PCT/CA96/00264
;; PRIOR FILING DATE: 1996-04-29
;; NUMBER OF SEQ ID NOS: 10
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 4
;; LENGTH: 1833
;; TYPE: PRT
;; ORGANISM: Moraxella catarrhalis
US-08-945-567D-4

Query Match 0.8%; Score 8; DB 4; Length 1833;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 836 VDADGNKV 843
Db 1178 VDADGNKV 1185

RESULT 6
US-08-621-944A-3
;; Sequence 3, Application US/08621944A
;; Patent No. 6440425
;; GENERAL INFORMATION:
;; APPLICANT: SASAKI, Ken
;; APPLICANT: HARKNESS, Robin E.
;; APPLICANT: LOOSMORE, Sheena M.
;; APPLICANT: KLEIN, Michel H.
;; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER
;; TITLE OF INVENTION: MEMBRANE PROTEIN OF MORAXELLA
;; NUMBER OF SEQUENCES: 10
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Sim & McBurney
;; STREET: Suite 701, 330 University Avenue
;; CITY: Toronto
;; STATE: Ontario
;; COUNTRY: Canada
;; ZIP: M5G 1R7
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/621.944A
;; FILING DATE: 26-MAR-1996
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/478,370
;; FILING DATE: 07-JUN-1995
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Stewart, Michael I.
;; REGISTRATION NUMBER: 24,973
;; REFERENCE/DOCKET NUMBER: 1038-587
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (416) 595-1155
;; TELEFAX: (416) 595-1163
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1992 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
US-08-621-944A-3

Query Match 0.8%; Score 8; DB 4; Length 1992;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 836 VDADGNKV 843
Db 1337 VDADGNKV 1344

RESULT 7
US-08-945-567D-3
;; Sequence 3, Application US/08945567D
;; Patent No. 6448386
;; GENERAL INFORMATION:
;; APPLICANT: SASAKI, Ken
;; APPLICANT: HARKNESS, Robin E.
;; APPLICANT: LOOSMORE, Sheena M.
;; APPLICANT: CHONG, Pele
;; APPLICANT: KLEIN, Michel H.
;; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE PROTEIN OF
;; TITLE OF INVENTION: MORAXELLA
;; FILE REFERENCE: 1038-745 MIS
;; CURRENT APPLICATION NUMBER: US/08/945,567D
;; CURRENT FILING DATE: 1996-04-29
;; PRIOR APPLICATION NUMBER: 08/431,718
;; PRIOR FILING DATE: 1995-05-01
;; PRIOR APPLICATION NUMBER: 08/478,370
;; PRIOR FILING DATE: 1995-06-07
;; PRIOR APPLICATION NUMBER: 08/621,944
;; PRIOR FILING DATE: 1996-03-26
;; PRIOR APPLICATION NUMBER: PCT/CA96/00264
;; PRIOR FILING DATE: 1996-04-29
;; NUMBER OF SEQ ID NOS: 10
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 3
;; LENGTH: 1992
;; TYPE: PRT
;; ORGANISM: Moraxella catarrhalis
US-08-945-567D-3

Query Match 0.8%; Score 8; DB 4; Length 1992;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 836 VDADGNKV 843
Db 1337 VDADGNKV 1344

RESULT 8
US-08-968-685A-10
;; Sequence 10, Application US/08968685A
;; Patent No. 6214981
;; GENERAL INFORMATION:
;; APPLICANT: TUCKER, KENNETH
;; APPLICANT: FLOSILA, LAURA
;; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
;; TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
;; NUMBER OF SEQUENCES: 21
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: PENNIE & EDMONDS LLP
;; STREET: 1155 Avenue of the Americas
;; CITY: New York
;; STATE: New York
;; COUNTRY: USA
;; ZIP: 10036-2711
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/968,685A
;; FILING DATE: No. 6214981ember 12, 1997
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Baldwin, Geraldine F.
;; REGISTRATION NUMBER: 31,232
;; REFERENCE/DOCKET NUMBER: 7969-060
;; TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2123 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-968-685A-10

Query Match 0.8%; Score 8; DB 3; Length 2123;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 836 VDADGNKV 843
Db 1468 VDADGNKV 1475

RESULT 9
US-09-268-347-49
Sequence 49, Application US/09268347
Patent No. 6335182
GENERAL INFORMATION:
APPLICANT: Loesmore, Sheena M.
TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
FILE REFERENCE: 1038-860
CURRENT APPLICATION NUMBER: US/09/268,347
CURRENT FILING DATE: 1999-03-16
NUMBER OF SEQ ID NOS: 54
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 49
LENGTH: 2314
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-09-268-347-49

Query Match 0.8%; Score 8; DB 4; Length 2314;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 836 VDADGNKV 843
Db 1333 VDADGNKV 1340

RESULT 10
5252328-2
Patent No. 5252328
APPLICANT: FAULDS, DARYL; VISHOOT, MIMI; BROOKS, EMILY
TITLE OF INVENTION: MYCOPLASMA HYOPNEUMONIAE ANTIGEN AND USES
THEREFORE
NUMBER OF SEQUENCES: 15
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/335,726
FILING DATE: 07-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 30,130
FILING DATE: 26-MAR-1987
SEQ ID NO: 2;
LENGTH: 10
5252328-2

Query Match 0.7%; Score 7; DB 6; Length 10;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 LIGANVV 44
Db 3 LIGANVV 9

RESULT 11
US-08-737-629-6
Sequence 6, Application US/08737629
Patent No. 6190886
GENERAL INFORMATION:
APPLICANT: Hoppe, Hans-Jurgen
TITLE OF INVENTION: Trimerising polypeptides, their manufacture
TITLE OF INVENTION: and use.
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vanderhye PC
STREET: 8th Floor, 1100 No. 6190886th Glebe Road
CITY: Arlington
STATE: Virginia
COUNTRY: United States of America
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/737,629
FILING DATE: 10-JAN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/01104
FILING DATE: 16-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9409768.0
FILING DATE: 16-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ms Mary J Wilson
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 620-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 41 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-737-629-6

Query Match 0.7%; Score 7; DB 3; Length 41;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 141 ALQGQVA 147
Db 10 ALQGQVA 16

RESULT 12
US-08-686-968C-2
Sequence 2, Application US/08686968C
Patent No. 6221361
GENERAL INFORMATION:
APPLICANT: Cochran, Mark D.
TITLE OF INVENTION: Recombinant Swinepox Virus
FILE REFERENCE: 39119-H/JML
CURRENT APPLICATION NUMBER: US/08/686,968C
CURRENT FILING DATE: 1996-07-25
NUMBER OF SEQ ID NOS: 231
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 77
TYPE: PRT
ORGANISM: Swinepox virus

US-08-686-968C-2

Query Match 0.7%; Score 7; DB 3; Length 77;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 946 FLRLKNL 952
DB 8 FLRLKNL 14

RESULT 13

US-08-285-440-27
; Sequence 27, Application US/08285440
; Patent No. 5532337
; GENERAL INFORMATION:
; APPLICANT: Ken'ichiro HAYASHI et al.
; TITLE OF INVENTION: POLYPEPTIDE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: DisplayWrite
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/285,440
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/858,947
; FILING DATE: March 27, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELEPHONE: 202-371-8850
; TELEFAX:

INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL:
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM:

STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL LINE:
; ORGANELLE:
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:

FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; PUBLICATION INFORMATION:
; AUTHORS:
; TITLE:
; JOURNAL:
; VOLUME:
; ISSUE:
; PAGES:
; DATE:
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO:

US-08-285-440-27

Query Match 0.7%; Score 7; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 377 SRINEWL 383
DB 59 SRINEWL 65

RESULT 14

US-08-630-349-27
; Sequence 27, Application US/08630349
; Patent No. 5739008
; GENERAL INFORMATION:
; APPLICANT: Ken'ichiro HAYASHI et al.
; TITLE OF INVENTION: POLYPEPTIDE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 144 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,349
; FILING DATE: April 10, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/285,440
; FILING DATE: August 4, 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/858,947
; FILING DATE: March 27, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-630-349-27

Query Match 0.7%; Score 7; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 377 SRINWL 383
 Db 59 SRINWL 65

RESULT 15

US-08-285-440-25
 ; Sequence 25, Application US/08285440
 ; Patent No. 5532337
 ; GENERAL INFORMATION:
 ; APPLICANT: Ken'ichiro HAYASHI et al.
 ; TITLE OF INVENTION: POLYPEPTIDE
 ; NUMBER OF SEQUENCES: 32
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Wenderoth, Lind & Ponack
 ; STREET: 805 Fifteenth Street, N.W., #700
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: DisplayWrite
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/285,440
 ; FILING DATE:
 ; CLASSIFICATION: 536
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/858,947
 ; FILING DATE: March 27, 1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Warren M. Cheek, Jr.
 ; REGISTRATION NUMBER: 33,367
 ; REFERENCE/DOCKET NUMBER:
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-371-8850
 ; TELEFAX:
 ; TELEX:

INFORMATION FOR SEQ ID NO: 25:

SEQUENCE CHARACTERISTICS:
 LENGTH: 90 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 HYPOTHETICAL:
 ANTI-SENSE:
 FRAGMENT TYPE:
 ORIGINAL SOURCE:
 ORGANISM:
 STRAIN:
 INDIVIDUAL ISOLATE:
 DEVELOPMENTAL STAGE:
 HAPLOTYPE:
 TISSUE TYPE:
 CELL TYPE:
 CELL LINE:
 ORGANELLE:
 IMMEDIATE SOURCE:
 LIBRARY:
 CLONE:
 POSITION IN GENOME:
 CHROMOSOME/SEGMENT:
 MAP POSITION:
 UNITS:
 FEATURE:
 NAME/KEY:
 LOCATION:
 IDENTIFICATION METHOD:
 OTHER INFORMATION:

PUBLICATION INFORMATION:

; AUTHORS:
 ; TITLE:
 ; JOURNAL:
 ; VOLUME:
 ; ISSUE:
 ; PAGES:
 ; DATE:
 ; DOCUMENT NUMBER:
 ; FILING DATE:
 ; PUBLICATION DATE:
 ; RELEVANT RESIDUES IN SEQ ID NO:
 ; US-08-285-440-25

Query Match 0.7%; Score 7; DB 1; Length 90;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 377 SRINWL 383
 Db 61 SRINWL 67

Search completed: January 7, 2004, 19:13:16
 Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2004, 19:12:09 ; Search time 38 Seconds
(without alignments)
5393.037 Million cell updates/sec

Title: US-09-581-286A-424
Perfect score: 1017
Sequence: 1 MKRWTLFLCLLTSGWAMA.....VKNQYNPSKQYVAGIQLSF 1017

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 747907 seqs, 201509753 residues

Word size : 0

Total number of hits satisfying chosen parameters: 747907

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09D_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	0.8	154	12	US-10-029-386-32028
2	8	0.8	340	12	US-10-369-493-1137
3	8	0.8	343	9	US-09-815-242-10474
4	8	0.8	546	15	US-10-128-714-3384
5	8	0.8	689	15	US-10-128-714-3384
6	8	0.8	1833	12	US-10-175-275-4
7	8	0.8	1833	12	US-10-175-282-4
8	8	0.8	1992	12	US-10-175-275-3
9	8	0.8	1992	12	US-10-175-282-3
10	8	0.8	2122	10	US-09-813-214A-9
11	7	0.7	18	12	US-10-082-014-104
12	7	0.7	30	12	US-10-372-076-105
13	7	0.7	18	12	US-10-029-386-32050
14	7	0.7	45	9	US-09-789-561-121
15	7	0.7	94	9	US-09-864-761-40529

16	7	0.7	103	9	US-09-864-761-41341	Sequence 41341, A
17	7	0.7	106	15	US-10-229-394-1	Sequence 1, Appli
18	7	0.7	111	9	US-09-939-980-437	Sequence 437, App
19	7	0.7	116	12	US-10-291-851-42	Sequence 42, Appl
20	7	0.7	137	12	US-10-241-220-69	Sequence 69, Appl
21	7	0.7	147	15	US-10-101-464A-701	Sequence 701, App
22	7	0.7	168	10	US-09-479-040-7	Sequence 7, Appli
23	7	0.7	169	9	US-09-815-242-5311	Sequence 5311, Ap
24	7	0.7	171	12	US-10-369-493-16949	Sequence 16949, A
25	7	0.7	179	9	US-09-815-242-12546	Sequence 12546, A
26	7	0.7	179	9	US-09-815-242-13069	Sequence 13069, A
27	7	0.7	189	15	US-10-156-761-13665	Sequence 13665, A
28	7	0.7	191	15	US-10-156-761-11564	Sequence 11564, A
29	7	0.7	243	12	US-10-104-047-2216	Sequence 2216, Ap
30	7	0.7	243	15	US-10-156-761-13715	Sequence 13715, A
31	7	0.7	255	12	US-10-290-438-2	Sequence 2, Appli
32	7	0.7	260	9	US-09-815-242-5398	Sequence 5398, Ap
33	7	0.7	262	10	US-09-738-626-4861	Sequence 4861, Ap
34	7	0.7	264	12	US-10-138-701-44	Sequence 44, Appl
35	7	0.7	267	11	US-09-769-787-114	Sequence 114, App
36	7	0.7	271	9	US-09-815-242-12298	Sequence 12298, A
37	7	0.7	271	9	US-09-815-242-12874	Sequence 12874, A
38	7	0.7	271	9	US-09-815-242-13143	Sequence 13143, A
39	7	0.7	280	12	US-10-017-161-386	Sequence 386, App
40	7	0.7	280	12	US-10-292-798-344	Sequence 344, App
41	7	0.7	290	12	US-10-046-671B-80	Sequence 80, Appl
42	7	0.7	302	15	US-10-156-761-10885	Sequence 10885, A
43	7	0.7	303	15	US-10-156-761-11434	Sequence 11434, A
44	7	0.7	312	10	US-09-738-626-6521	Sequence 6521, Ap
45	7	0.7	313	12	US-10-292-798-592	Sequence 592, App

ALIGNMENTS

RESULT 1
US-10-029-386-32028
; Sequence 32028, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32028
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF025422.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
; OTHER INFORMATION: SWISSPROT HIT: P10241, EVALU8 8.10e-01
US-10-029-386-32028

Query Match 0.8%; Score 8; DB 12; Length 154;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 158 PTAVASVE 165
Db 92 PTAVASVE 99

RESULT 2
US-10-369-493-1137
; Sequence 1137, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1137
; LENGTH: 340
; TYPE: PRT
; ORGANISM: Methanobacterium thermoautotrophicum
US-10-369-493-1137

Query Match 0.8%; Score 8; DB 12; Length 340;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 260 MTGDELLD 267
|||||||
DB 79 MTGDELLD 86

RESULT 3
US-09-815-242-10474
; Sequence 10474, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haseelbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10474
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Escherichia coli

US-09-815-242-10474

Query Match 0.8%; Score 8; DB 9; Length 343;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 DGMQTSLD 190
|||||||
DB 273 DGMQTSLD 280

RESULT 4
US-10-128-714-3384
; Sequence 3384, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3384
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-3384

Query Match 0.8%; Score 8; DB 15; Length 546;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 LMGHEYIE 547
|||||||
DB 326 LMGHEYIE 333

RESULT 5
US-10-128-714-8384
; Sequence 8384, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066

; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8384
; LENGTH: 689
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-8384

Query Match 0.8%; Score 8; DB 15; Length 689;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 LMGHEYIE 547
| | | | |
DB 433 LMGHEYIE 440

RESULT 6

US-10-175-275-4
; Sequence 4, Application US/10175275
; Publication No. US20030171254A1
; GENERAL INFORMATION:

; APPLICANT: SASAKI, Ken
; APPLICANT: HARKNESS, Robin E.

; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: CHONG, Pele

; APPLICANT: KLEIN, Michel H.

; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE PROTEIN OF MORAXELLA

; FILE REFERENCE: 1038-1235 MIS

; CURRENT APPLICATION NUMBER: US/10/175,275
; CURRENT FILING DATE: 2002-06-20

; PRIOR APPLICATION NUMBER: 08/945,567

; PRIOR FILING DATE: 1998-03-19

; PRIOR APPLICATION NUMBER: 08/431,718

; PRIOR FILING DATE: 1995-05-01

; PRIOR APPLICATION NUMBER: 08/478,370

; PRIOR FILING DATE: 1995-06-07

; PRIOR APPLICATION NUMBER: 08/621,944

; PRIOR FILING DATE: 1996-03-26

; PRIOR APPLICATION NUMBER: PCT/CA96/00264

; PRIOR FILING DATE: 1996-04-29

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 4

; LENGTH: 1833

; TYPE: PRT

; ORGANISM: Moraxella catarrhalis

US-10-175-275-4

Query Match 0.8%; Score 8; DB 12; Length 1833;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 836 VDADGNKV 843
| | | | |
DB 1178 VDADGNKV 1185

RESULT 7

US-10-175-282-4

; Sequence 4, Application US/10175282

; Publication No. US20030170657A1

; GENERAL INFORMATION:

; APPLICANT: SASAKI, Ken

; APPLICANT: HARKNESS, Robin E.

; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: CHONG, Pele
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-1234 MIS
; CURRENT APPLICATION NUMBER: US/10/175,282
; CURRENT FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: 08/945,567
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 08/431,718
; PRIOR FILING DATE: 1995-05-01
; PRIOR APPLICATION NUMBER: 08/478,370
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 08/621,944
; PRIOR FILING DATE: 1996-03-26
; PRIOR APPLICATION NUMBER: PCT/CA96/00264
; PRIOR FILING DATE: 1996-04-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1833
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-10-175-282-4

Query Match 0.8%; Score 8; DB 12; Length 1833;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 836 VDADGNKV 843
| | | | |
DB 1178 VDADGNKV 1185

RESULT 8

US-10-175-275-3

; Sequence 3, Application US/10175275

; Publication No. US20030171254A1

; GENERAL INFORMATION:

; APPLICANT: SASAKI, Ken

; APPLICANT: HARKNESS, Robin E.

; APPLICANT: LOOSMORE, Sheena M.

; APPLICANT: CHONG, Pele

; APPLICANT: KLEIN, Michel H.

; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE PROTEIN OF MORAXELLA

; FILE REFERENCE: 1038-1235 MIS

; CURRENT APPLICATION NUMBER: US/10/175,275

; CURRENT FILING DATE: 2002-06-20

; PRIOR APPLICATION NUMBER: 08/945,567

; PRIOR FILING DATE: 1998-03-19

; PRIOR APPLICATION NUMBER: 08/431,718

; PRIOR FILING DATE: 1995-05-01

; PRIOR APPLICATION NUMBER: 08/478,370

; PRIOR FILING DATE: 1995-06-07

; PRIOR APPLICATION NUMBER: 08/621,944

; PRIOR FILING DATE: 1996-03-26

; PRIOR APPLICATION NUMBER: PCT/CA96/00264

; PRIOR FILING DATE: 1996-04-29

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3

; LENGTH: 1992

; TYPE: PRT

; ORGANISM: Moraxella catarrhalis

US-10-175-275-3

Query Match 0.8%; Score 8; DB 12; Length 1992;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 836 VDADGNKV 843

Db 1337 VDADGNKV 1344
|||||

RESULT 9
US-10-175-282-3
; Sequence 3, Application US/10175282
; Publication No. US20030170657A1
; GENERAL INFORMATION:
; APPLICANT: SASAKI, Ken
; APPLICANT: HARKNESS, Robin E.
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: CHONG, Pele
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-1234 MIS
; CURRENT APPLICATION NUMBER: US/10/175,282
; CURRENT FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: 08/945,567
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 08/431,718
; PRIOR FILING DATE: 1995-05-01
; PRIOR APPLICATION NUMBER: 08/478,370
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 08/621,944
; PRIOR FILING DATE: 1996-03-26
; PRIOR APPLICATION NUMBER: PCT/CA96/00264
; PRIOR FILING DATE: 1996-04-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1992
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-10-175-282-3

Query Match 0.8%; Score 8; DB 12; Length 1992;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 836 VDADGNKV 843
|||||

Db 1337 VDADGNKV 1344
|||||

RESULT 10
US-09-813-214A-9
; Sequence 9, Application US/09813214A
; Patent No. US20020177200A1
; GENERAL INFORMATION:
; APPLICANT: Tucker, Kenneth
; APPLICANT: Plosilla, Laura
; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE PROTEIN-106 POLYPEPTIDE, GEN
; TITLE OF INVENTION: SEQUENCE AND USES THEREOF
; FILE REFERENCE: 7969-089-999
; CURRENT APPLICATION NUMBER: US/09/813,214A
; CURRENT FILING DATE: 2000-03-20
; PRIOR APPLICATION NUMBER: 08/968,685
; PRIOR FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 2122
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-813-214A-9

Query Match 0.8%; Score 8; DB 10; Length 2122;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 836 VDADGNKV 843

Db 1467 VDADGNKV 1474
|||||

RESULT 11
US-10-082-014-104
; Sequence 104, Application US/10082014
; Publication No. US20030185858A1
; GENERAL INFORMATION:
; APPLICANT: Birkett, Ashley J.
; TITLE OF INVENTION: IMMUNOGENIC HBC CHIMER PARTICLES STABILIZED WITH AN N-TERMINAL CYC
; FILE REFERENCE: ICC-130.0 4564/85124
; CURRENT APPLICATION NUMBER: US/10/082,014
; CURRENT FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: 09/930,915
; PRIOR FILING DATE: 2001-08-15
; NUMBER OF SEQ ID NOS: 290
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 104
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-10-082-014-104

Query Match 0.7%; Score 7; DB 12; Length 18;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 569 SQKGTGN 575
|||||

Db 5 SQKGTGN 11

RESULT 12
US-10-372-076-105
; Sequence 105, Application US/10372076
; Publication No. US20030198645A1
; GENERAL INFORMATION:
; APPLICANT: Friede, Martin
; APPLICANT: Page, Mark
; TITLE OF INVENTION: STABILIZED HBC CHIMER PARTICLES AS THERAPEUTIC VACCINE FOR
; TITLE OF INVENTION: CHRONIC HEPATITIS
; FILE REFERENCE: 4564/87179
; CURRENT APPLICATION NUMBER: US/10/372,076
; CURRENT FILING DATE: 2003-02-21
; PRIOR APPLICATION NUMBER: 10/080,299
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 10/082,014
; PRIOR FILING DATE: 2002-02-22
; NUMBER OF SEQ ID NOS: 308
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 105
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-10-372-076-105

Query Match 0.7%; Score 7; DB 12; Length 18;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 569 SQKGTGN 575
|||||

Db 5 SQKGTGN 11

RESULT 13
US-10-029-386-32505
; Sequence 32505, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.

```

; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR C
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: ACOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32505
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO A1132777.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2
; OTHER INFORMATION: SWISSPROT HIT: P28282, EVALUE 8.70e+00
; US-10-029-386-32505

Query Match          0.7%; Score 7; DB 12; Length 30;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 490 RTSSKRM 496
Db 5 RTSSKRM 11
|||||

RESULT 14
US-09-789-561-121
; Sequence 121, Application US/09789561
; Patent No. US20020064818A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043p1
; CURRENT APPLICATION NUMBER: US/09/789,561
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR FILING DATE: 1999-09-03
; PRIOR FILING DATE: 1999-09-03
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 121
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-789-561-121

Query Match          0.7%; Score 7; DB 9; Length 45;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 873 SWKGLSL 879
Db 27 SWKGLSL 33
|||||

RESULT 15
US-09-864-761-40529
; Sequence 40529, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.

```

```

; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40529
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO A1137225.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EST HUMAN HIT: W26876.1, EVALUE 5.40e-02
; OTHER INFORMATION: SWISSPROT HIT: P51989, EVALUE 2.00e-03
; US-09-864-761-40529

Query Match          0.7%; Score 7; DB 9; Length 94;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 168 GTGSLGA 174
Db 71 GTGSLGA 77
|||||

Search completed: January 7, 2004, 19:17:57
Job time : 40 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2004, 19:11:09 ; Search time 181 Seconds
(without alignments)
5112.638 Million cell updates/sec

Title: US-09-581-286A-424
Perfect score: 1017
Sequence: 1 MKRWTLFLCLLTSGWAMA.....VKNQYPNSQYVAGIQLSF 1017

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 5728757 seqs, 909918778 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents AA_Main:**

- 1: /cgn2_6/ptodata/1/paa/PCTUS COMB.pep:**
- 2: /cgn2_6/ptodata/1/paa/US06 COMB.pep:**
- 3: /cgn2_6/ptodata/1/paa/US07 COMB.pep:**
- 4: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 5: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 6: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 7: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 8: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 9: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 10: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 11: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 12: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 13: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 14: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 15: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 16: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 17: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 18: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 19: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 20: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 21: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 22: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 23: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 24: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 25: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 26: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 27: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 28: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 29: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 30: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 31: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**
- 32: /cgn2_6/ptodata/1/paa/US08 COMB.pep:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	15	1.5	246	19 US-09-540-209B-7259	Sequence 7259, Ap

2	15	1.5	1110	28	US-10-282-122A-48782	Sequence 48782, A
3	10	1.0	602	19	US-09-540-209B-5552	Sequence 5552, Ap
4	10	1.0	1008	19	US-09-540-209B-8074	Sequence 8074, Ap
5	10	1.0	1032	19	US-09-540-209B-10074	Sequence 10074, A
6	10	1.0	1119	19	US-09-540-209B-9958	Sequence 9958, Ap
7	9	0.9	100	1	PCT-US01-01354-13978	Sequence 13978, A
8	9	0.9	100	22	US-09-764-905-13978	Sequence 13978, A
9	9	0.9	100	26	US-10-092-399-13978	Sequence 13978, A
10	9	0.9	160	30	US-10-424-599-156192	Sequence 156192, A
11	8	0.8	137	30	US-10-424-599-156192	Sequence 156192, A
12	8	0.8	154	26	US-10-029-386-32028	Sequence 32028, A
13	8	0.8	159	28	US-10-282-122A-63343	Sequence 63343, A
14	8	0.8	171	19	US-09-540-236-2189	Sequence 2189, Ap
15	8	0.8	171	31	US-10-603-108-2189	Sequence 2189, Ap
16	8	0.8	171	32	US-60-128-476-3662	Sequence 3662, Ap
17	8	0.8	175	22	US-09-791-537-40460	Sequence 40460, A
18	8	0.8	207	20	US-09-675-784A-12029	Sequence 12029, A
19	8	0.8	267	19	US-09-540-209B-7867	Sequence 7867, Ap
20	8	0.8	295	30	US-10-431-652-6656	Sequence 6656, Ap
21	8	0.8	296	18	US-09-417-507-22594	Sequence 22594, A
22	8	0.8	299	22	US-09-791-537-150930	Sequence 150930, A
23	8	0.8	321	22	US-09-791-537-96830	Sequence 96830, A
24	8	0.8	338	30	US-10-417-884-5819	Sequence 5819, Ap
25	8	0.8	340	22	US-09-791-537-2856	Sequence 2856, Ap
26	8	0.8	340	29	US-10-369-493-1137	Sequence 1137, Ap
27	8	0.8	340	32	US-60-360-039-1137	Sequence 1137, Ap
28	8	0.8	342	28	US-10-282-122A-76303	Sequence 76303, A
29	8	0.8	343	1	PCT-US02-03987-10474	Sequence 10474, A
30	8	0.8	343	22	US-09-791-537-10065	Sequence 10065, A
31	8	0.8	343	23	US-09-815-242-10474	Sequence 10474, A
32	8	0.8	343	26	US-10-072-851-10474	Sequence 10474, A
33	8	0.8	343	28	US-10-282-122A-43282	Sequence 43282, A
34	8	0.8	389	16	US-09-248-796-17443	Sequence 17443, A
35	8	0.8	389	23	US-09-897-516-5626	Sequence 5626, Ap
36	8	0.8	389	27	US-10-179-131-8150	Sequence 8150, Ap
37	8	0.8	389	31	US-10-603-113-17443	Sequence 17443, A
38	8	0.8	389	32	US-60-096-409-17443	Sequence 17443, A
39	8	0.8	389	32	US-60-215-161-5626	Sequence 5626, Ap
40	8	0.8	428	18	US-09-417-507-34630	Sequence 34630, A
41	8	0.8	476	1	PCT-US01-08631-39099	Sequence 39099, A
42	8	0.8	487	19	US-09-583-110-4658	Sequence 4658, Ap
43	8	0.8	487	31	US-10-640-833-4658	Sequence 4658, Ap
44	8	0.8	490	15	US-09-107-433-4640	Sequence 4640, Ap
45	8	0.8	490	31	US-10-617-320-4640	Sequence 4640, Ap

ALIGNMENTS

RESULT 1
US-09-540-209B-7259
; Sequence 7259, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRATRIUM
; FILE REFERENCE: 2709 1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 7259
; LENGTH: 246
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-7259

Query Match 1.5%; Score 15; DB 19; Length 246;
Best Local Similarity 100.0%; Pred. No. 2.6e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	205	VKQASATSIYGARA	219
Db	47	VKQASATSIYGARA	61

```
RESULT 2
US-10-282-122A-48782
; Sequence 48782, Application US/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48782
; LENGTH: 1110
; TYPE: PRT
; ORGANISM: Bacteroides fragilis
US-10-282-122A-48782

Query Match          1.5%; Score 15; DB 28; Length 1110;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      205 VLKDSATSIIYGARA 219
Db      212 VLKDSATSIIYGARA 226

RESULT 3
US-09-540-209B-5552
; Sequence 5552, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRAGILIS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 5552
; LENGTH: 602
; TYPE: PRT
; ORGANISM: Bacteroides fragilis
US-09-540-209B-5552

Query Match          1.0%; Score 10; DB 19; Length 1032;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      214 IYGARAANGV 223
Db      243 IYGARAANGV 252

RESULT 4
US-09-540-209B-8074
; Sequence 8074, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRAGILIS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 8074
; LENGTH: 1008
; TYPE: PRT
; ORGANISM: Bacteroides fragilis
US-09-540-209B-8074

Query Match          1.0%; Score 10; DB 19; Length 1008;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      214 IYGARAANGV 223
Db      244 IYGARAANGV 253

RESULT 5
US-09-540-209B-10074
; Sequence 10074, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRAGILIS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 10074
; LENGTH: 1032
; TYPE: PRT
; ORGANISM: Bacteroides fragilis
US-09-540-209B-10074

Query Match          1.0%; Score 10; DB 19; Length 1032;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      214 IYGARAANGV 223
Db      243 IYGARAANGV 252

RESULT 6
US-09-540-209B-9958
; Sequence 9958, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRAGILIS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 5552
; LENGTH: 602
; TYPE: PRT
; ORGANISM: Bacteroides fragilis
US-09-540-209B-9958

Query Match          1.0%; Score 10; DB 19; Length 1032;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      214 IYGARAANGV 223
Db      243 IYGARAANGV 252
```

```
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-5552

Query Match          1.0%; Score 10; DB 19; Length 602;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      214 IYGARAANGV 223
Db      244 IYGARAANGV 253

RESULT 4
US-09-540-209B-8074
; Sequence 8074, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRAGILIS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 8074
; LENGTH: 1008
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-8074

Query Match          1.0%; Score 10; DB 19; Length 1008;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      214 IYGARAANGV 223
Db      244 IYGARAANGV 253

RESULT 5
US-09-540-209B-10074
; Sequence 10074, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRAGILIS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 10074
; LENGTH: 1032
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-10074

Query Match          1.0%; Score 10; DB 19; Length 1032;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      214 IYGARAANGV 223
Db      243 IYGARAANGV 252

RESULT 6
US-09-540-209B-9958
; Sequence 9958, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRAGILIS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 5552
; LENGTH: 602
; TYPE: PRT
; ORGANISM: Bacteroides fragilis
US-09-540-209B-9958

Query Match          1.0%; Score 10; DB 19; Length 1032;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      214 IYGARAANGV 223
Db      243 IYGARAANGV 252
```


; CURRENT APPLICATION NUMBER: US/09/540,209B
 ; CURRENT FILING DATE: 2000-04-04
 ; NUMBER OF SEQ ID NOS: 10444
 ; SEQ ID NO 9958
 ; LENGTH: 1119
 ; TYPE: PRT
 ; ORGANISM: B.fragilis
 US-09-540-209B-9958

Query Match 1.0%; Score 10; DB 19; Length 1119;

Best Local Similarity 100.0%; Pred. No. 12; Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 214 IYGARAANGV 223
 |||||
 Db 244 IYGARAANGV 253

RESULT 7

PCT-US01-01354-13978
 ; Sequence 13978, Application PC/TUS0101354
 ; GENERAL INFORMATION:
 ; APPLICANT: Human Genome Sciences, Inc. et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC004PCT
 ; CURRENT APPLICATION NUMBER: PCT/US01/01354
 ; CURRENT FILING DATE: 2001-03-17
 ; NUMBER OF SEQ ID NOS: 42506
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 13978
 ; LENGTH: 100
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (96)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 PCT-US01-01354-13978

Query Match 0.9%; Score 9; DB 1; Length 100;

Best Local Similarity 100.0%; Pred. No. 12; Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 226 IOTKKGKMS 234
 |||||
 Db 2 IOTKKGKMS 10

RESULT 8

US-09-764-905-13978
 ; Sequence 13978, Application US/09764905
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC004
 ; CURRENT APPLICATION NUMBER: US/09/764,905
 ; CURRENT FILING DATE: 2001-01-17
 ; PRIOR APPLICATION NUMBER: 60/179,065
 ; PRIOR FILING DATE: 2000-01-31
 ; PRIOR APPLICATION NUMBER: 60/180,628
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: 60/214,886
 ; PRIOR FILING DATE: 2000-06-28
 ; PRIOR APPLICATION NUMBER: 60/217,487
 ; PRIOR FILING DATE: 2000-07-11
 ; PRIOR APPLICATION NUMBER: 60/225,758
 ; PRIOR FILING DATE: 2000-08-14
 ; PRIOR APPLICATION NUMBER: 60/220,963
 ; PRIOR FILING DATE: 2000-07-26
 ; PRIOR APPLICATION NUMBER: 60/217,496
 ; PRIOR FILING DATE: 2000-07-11
 ; PRIOR APPLICATION NUMBER: 60/225,447
 ; PRIOR FILING DATE: 2000-08-14

; PRIOR APPLICATION NUMBER: 60/218,290
 ; PRIOR FILING DATE: 2000-07-14
 ; PRIOR APPLICATION NUMBER: 60/225,757
 ; PRIOR FILING DATE: 2000-08-14
 ; PRIOR APPLICATION NUMBER: 60/226,868
 ; PRIOR FILING DATE: 2000-08-22
 ; PRIOR APPLICATION NUMBER: 60/216,647
 ; PRIOR FILING DATE: 2000-07-07
 ; PRIOR APPLICATION NUMBER: 60/225,267
 ; PRIOR FILING DATE: 2000-08-14
 ; PRIOR APPLICATION NUMBER: 60/216,880
 ; PRIOR FILING DATE: 2000-07-07
 ; PRIOR APPLICATION NUMBER: 60/225,270
 ; PRIOR FILING DATE: 2000-08-14
 ; PRIOR APPLICATION NUMBER: 60/251,869
 ; PRIOR FILING DATE: 2000-12-08
 ; PRIOR APPLICATION NUMBER: 60/235,834
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: 60/234,274
 ; PRIOR FILING DATE: 2000-09-21
 ; PRIOR APPLICATION NUMBER: 60/234,223
 ; PRIOR FILING DATE: 2000-09-21
 ; PRIOR APPLICATION NUMBER: 60/228,924
 ; PRIOR FILING DATE: 2000-08-30
 ; PRIOR APPLICATION NUMBER: 60/224,518
 ; PRIOR FILING DATE: 2000-08-14
 ; PRIOR APPLICATION NUMBER: 60/236,369
 ; PRIOR FILING DATE: 2000-09-29
 ; PRIOR APPLICATION NUMBER: 60/224,519
 ; PRIOR FILING DATE: 2000-08-14
 ; PRIOR APPLICATION NUMBER: 60/220,964
 ; PRIOR FILING DATE: 2000-07-26
 ; PRIOR APPLICATION NUMBER: 60/241,809
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/249,299
 ; PRIOR FILING DATE: 2000-11-17
 ; PRIOR APPLICATION NUMBER: 60/236,327
 ; PRIOR FILING DATE: 2000-09-29
 ; PRIOR APPLICATION NUMBER: 60/241,785
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/244,617
 ; PRIOR FILING DATE: 2000-11-01
 ; PRIOR APPLICATION NUMBER: 60/225,268
 ; PRIOR FILING DATE: 2000-08-14
 ; PRIOR APPLICATION NUMBER: 60/236,368
 ; PRIOR FILING DATE: 2000-09-29
 ; PRIOR APPLICATION NUMBER: 60/251,856
 ; PRIOR FILING DATE: 2000-12-08
 ; PRIOR APPLICATION NUMBER: 60/251,868
 ; PRIOR FILING DATE: 2000-12-08
 ; PRIOR APPLICATION NUMBER: 60/229,344
 ; PRIOR FILING DATE: 2000-09-01
 ; PRIOR APPLICATION NUMBER: 60/234,997
 ; PRIOR FILING DATE: 2000-09-25
 ; PRIOR APPLICATION NUMBER: 60/229,343
 ; PRIOR FILING DATE: 2000-09-01
 ; PRIOR APPLICATION NUMBER: 60/229,345
 ; PRIOR FILING DATE: 2000-09-01
 ; PRIOR APPLICATION NUMBER: 60/229,287
 ; PRIOR FILING DATE: 2000-09-01
 ; PRIOR APPLICATION NUMBER: 60/229,513
 ; PRIOR FILING DATE: 2000-09-05
 ; PRIOR APPLICATION NUMBER: 60/231,413
 ; PRIOR FILING DATE: 2000-09-08
 ; PRIOR APPLICATION NUMBER: 60/229,509
 ; PRIOR FILING DATE: 2000-09-05
 ; PRIOR APPLICATION NUMBER: 60/236,367
 ; PRIOR FILING DATE: 2000-09-29
 ; PRIOR APPLICATION NUMBER: 60/237,039
 ; PRIOR FILING DATE: 2000-10-02
 ; PRIOR APPLICATION NUMBER: 60/237,038
 ; PRIOR FILING DATE: 2000-10-02
 ; PRIOR APPLICATION NUMBER: 60/236,370

; PRIOR FILING DATE: 2000-09-29
 ; PRIOR APPLICATION NUMBER: 60/236,802
 ; PRIOR FILING DATE: 2000-10-02
 ; PRIOR APPLICATION NUMBER: 60/237,037
 ; PRIOR FILING DATE: 2000-10-02
 ; PRIOR APPLICATION NUMBER: 60/237,040
 ; PRIOR FILING DATE: 2000-10-02
 ; PRIOR APPLICATION NUMBER: 60/240,960
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/239,935
 ; PRIOR FILING DATE: 2000-10-13
 ; PRIOR APPLICATION NUMBER: 60/239,937
 ; PRIOR FILING DATE: 2000-10-13
 ; PRIOR APPLICATION NUMBER: 60/241,787
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/246,474
 ; PRIOR FILING DATE: 2000-11-08
 ; PRIOR APPLICATION NUMBER: 60/246,532
 ; PRIOR FILING DATE: 2000-11-08
 ; PRIOR APPLICATION NUMBER: 60/249,216
 ; PRIOR FILING DATE: 2000-11-17
 ; PRIOR APPLICATION NUMBER: 60/249,210
 ; PRIOR FILING DATE: 2000-11-17
 ; PRIOR APPLICATION NUMBER: 60/226,681
 ; PRIOR FILING DATE: 2000-08-22
 ; PRIOR APPLICATION NUMBER: 60/225,759
 ; PRIOR FILING DATE: 2000-08-14
 ; PRIOR APPLICATION NUMBER: 60/225,213
 ; PRIOR FILING DATE: 2000-08-14
 ; PRIOR APPLICATION NUMBER: 60/227,182
 ; PRIOR FILING DATE: 2000-08-22
 ; PRIOR APPLICATION NUMBER: 60/225,214
 ; PRIOR FILING DATE: 2000-08-14
 ; PRIOR APPLICATION NUMBER: 60/235,836
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: 60/230,438
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/215,135
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: 60/225,266
 ; PRIOR FILING DATE: 2000-08-14
 ; PRIOR APPLICATION NUMBER: 60/249,218
 ; PRIOR FILING DATE: 2000-11-17
 ; PRIOR APPLICATION NUMBER: 60/249,208
 ; PRIOR FILING DATE: 2000-11-17
 ; PRIOR APPLICATION NUMBER: 60/249,213
 ; PRIOR FILING DATE: 2000-11-17
 ; PRIOR APPLICATION NUMBER: 60/249,212
 ; PRIOR FILING DATE: 2000-11-17
 ; PRIOR APPLICATION NUMBER: 60/249,207
 ; PRIOR FILING DATE: 2000-11-17
 ; PRIOR APPLICATION NUMBER: 60/249,245
 ; PRIOR FILING DATE: 2000-11-17
 ; PRIOR APPLICATION NUMBER: 60/249,244
 ; PRIOR FILING DATE: 2000-11-17
 ; PRIOR APPLICATION NUMBER: 60/249,217
 ; PRIOR FILING DATE: 2000-11-17
 ; PRIOR APPLICATION NUMBER: 60/249,211
 ; PRIOR FILING DATE: 2000-11-17
 ; PRIOR APPLICATION NUMBER: 60/249,215
 ; PRIOR FILING DATE: 2000-11-17
 ; PRIOR APPLICATION NUMBER: 60/249,264
 ; PRIOR FILING DATE: 2000-11-17
 ; PRIOR APPLICATION NUMBER: 60/249,214
 ; PRIOR FILING DATE: 2000-11-17
 ; PRIOR APPLICATION NUMBER: 60/249,297
 ; PRIOR FILING DATE: 2000-11-17
 ; PRIOR APPLICATION NUMBER: 60/232,400
 ; PRIOR FILING DATE: 2000-09-14
 ; PRIOR APPLICATION NUMBER: 60/231,242
 ; PRIOR FILING DATE: 2000-09-08
 ; PRIOR APPLICATION NUMBER: 60/232,081
 ; PRIOR FILING DATE: 2000-09-08

; PRIOR APPLICATION NUMBER: 60/232,080
 ; PRIOR FILING DATE: 2000-09-08
 ; PRIOR APPLICATION NUMBER: 60/231,414
 ; PRIOR FILING DATE: 2000-09-08
 ; PRIOR APPLICATION NUMBER: 60/231,244
 ; PRIOR FILING DATE: 2000-09-08
 ; PRIOR APPLICATION NUMBER: 60/233,064
 ; PRIOR FILING DATE: 2000-09-14
 ; PRIOR APPLICATION NUMBER: 60/233,063
 ; PRIOR FILING DATE: 2000-09-14
 ; PRIOR APPLICATION NUMBER: 60/232,397
 ; PRIOR FILING DATE: 2000-09-14
 ; PRIOR APPLICATION NUMBER: 60/232,399
 ; PRIOR FILING DATE: 2000-09-14
 ; PRIOR APPLICATION NUMBER: 60/232,401
 ; PRIOR FILING DATE: 2000-09-14
 ; PRIOR APPLICATION NUMBER: 60/241,808
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/241,826
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/241,786
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/241,221
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/246,475
 ; PRIOR FILING DATE: 2000-11-08
 ; PRIOR APPLICATION NUMBER: 60/231,243
 ; PRIOR FILING DATE: 2000-09-08
 ; PRIOR APPLICATION NUMBER: 60/233,065
 ; PRIOR FILING DATE: 2000-09-14
 ; PRIOR APPLICATION NUMBER: 60/232,398

Query Match 0.9%; Score 9; DB 22; Length 100;

Best Local Similarity 100.0%; Pred. No. 12;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 226 IOTKKGMS 234
 |||||
 Db 2 IOTKKGMS 10

RESULT 9

US-10-092-399-13978
 ; Sequence 13978, Application US/10092399
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC004C1
 ; CURRENT APPLICATION NUMBER: US/10/092,399
 ; CURRENT FILING DATE: 2002-03-07
 ; NUMBER OF SEQ ID NOS: 42506
 ; Prior Application removed - See File Wrapper or Palm
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 13978
 ; LENGTH: 100
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (96)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-10-092-399-13978

Query Match 0.9%; Score 9; DB 26; Length 100;

Best Local Similarity 100.0%; Pred. No. 12;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 226 IOTKKGMS 234
 |||||
 Db 2 IOTKKGMS 10

RESULT 10

Job time : 183 secs

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 63343
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
; US-10-282-122A-63343

Query Match 0.8%; Score 8; DB 28; Length 159;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 817 EYAGIDKK 824
| | | | |
Db 137 EYAGIDKK 144

RESULT 14

US-09-540-236-2189
; Sequence 2189, Application US/09540236
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2189
; LENGTH: 171
; TYPE: PRT
; ORGANISM: M.catarrhalis
; US-09-540-236-2189

Query Match 0.8%; Score 8; DB 19; Length 171;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 817 EYAGIDKK 824
| | | | |
Db 149 EYAGIDKK 156

RESULT 15

US-10-603-108-2189
; Sequence 2189, Application US/10603108
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; FILE REFERENCE: PATH03-14
; CURRENT APPLICATION NUMBER: US/10/603,108
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: US 09/540,263
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 60/125,416
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2189
; LENGTH: 171
; TYPE: PRT
; ORGANISM: M.catarrhalis
; US-10-603-108-2189

Query Match 0.8%; Score 8; DB 31; Length 171;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 817 EYAGIDKK 824
| | | | |
Db 149 EYAGIDKK 156

Search completed: January 7, 2004, 19:17:06

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2004, 19:10:39 ; Search time 22 Seconds
(without alignments)
3019.523 Million cell updates/sec

Title: US-09-581-286A-424
Perfect score: 1017
Sequence: 1 MKRWTFLLCLLTSGWAMA.....VKNQYPNSKQVAGIQLSF 1017

Scoring table: OLIGO

Searched: 324163 seqs, 65319079 residues

Word size : 0

Total number of hits satisfying chosen parameters: 324163

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Pending Patents_AA_New.*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1017	100.0	1017	5	US-09-581-286A-424
2	1017	100.0	1046	5	US-09-581-286A-300
3	1014	99.7	1014	5	US-09-581-286A-425
4	8	0.8	389	5	US-09-897-516A-5632
5	8	0.8	487	6	US-10-472-928-880
6	8	0.8	1127	6	US-10-679-063-14822
7	7	0.7	18	1	PCT-US03-05499-28
8	7	0.7	18	6	US-10-677-074-105
9	7	0.7	37	7	US-60-519-270-109
10	7	0.7	117	6	US-10-425-114A-48932
11	7	0.7	127	6	US-10-679-063-18798
12	7	0.7	134	6	US-10-425-114A-45140
13	7	0.7	138	6	US-10-425-114A-69574
14	7	0.7	141	6	US-10-425-114A-57907
15	7	0.7	173	6	US-10-425-114A-67079
16	7	0.7	189	6	US-10-425-114A-38940
17	7	0.7	216	6	US-10-425-114A-49185
18	7	0.7	240	6	US-10-679-063-15096
19	7	0.7	240	6	US-10-679-063-21943
20	7	0.7	248	5	US-09-614-150A-27816
21	7	0.7	257	6	US-10-679-063-9844
22	7	0.7	267	6	US-10-472-928-1360
23	7	0.7	280	7	US-60-500-315-1192
24	7	0.7	293	6	US-10-425-114A-69548
25	7	0.7	312	5	US-09-614-150A-34302
26	7	0.7	314	5	US-09-614-150A-6471

27	7	0.7	330	6	US-10-418-861B-24	Sequence 24, Appl
28	7	0.7	343	6	US-10-425-114A-62971	Sequence 62971, A
29	7	0.7	351	5	US-09-614-150A-9978	Sequence 9978, Ap
30	7	0.7	361	6	US-10-662-136-23	Sequence 23, Appl
31	7	0.7	372	5	US-09-820-843B-8	Sequence 8, Appl
32	7	0.7	378	6	US-10-425-114A-65081	Sequence 65081, A
33	7	0.7	389	1	PCT-US03-24554-25	Sequence 25, Appl
34	7	0.7	454	1	PCT-US03-28227-5294	Sequence 5294, Ap
35	7	0.7	457	1	PCT-US03-28227-4435	Sequence 4435, Ap
36	7	0.7	457	1	PCT-US03-28227-4439	Sequence 4439, Ap
37	7	0.7	462	6	US-10-679-063-9028	Sequence 9028, Ap
38	7	0.7	469	6	US-10-417-700A-63	Sequence 63, Appl
39	7	0.7	471	6	US-10-679-063-15381	Sequence 15381, A
40	7	0.7	475	1	PCT-US03-28227-4436	Sequence 4436, Ap
41	7	0.7	478	7	US-60-485-450-1555	Sequence 1555, Ap
42	7	0.7	479	6	US-10-679-063-17942	Sequence 17942, A
43	7	0.7	482	1	PCT-US03-28227-4437	Sequence 4437, Ap
44	7	0.7	486	1	PCT-US03-28227-4434	Sequence 4434, Ap
45	7	0.7	493	6	US-10-679-063-10645	Sequence 10645, A

ALIGNMENTS

RESULT 1
US-09-581-286A-424
; Sequence 424, Application US/09581286A
; GENERAL INFORMATION:
; APPLICANT: ROSS, BRUCE C.
; APPLICANT: BARR, IAN G.
; APPLICANT: PATTERSON, MICHELLE A.
; APPLICANT: AGIUS, CATHERINE T.
; APPLICANT: ROTH, LINDA J.
; APPLICANT: MARGITTS, MAL B.
; APPLICANT: HOCKING, DIANNA M.
; APPLICANT: WEBB, ELIZABETH A.
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS POLYPEPTIDES AND NUCLEOTIDES
; FILE REFERENCE: 4137-3
; CURRENT APPLICATION NUMBER: US/09/581,286A
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: PCT/AU98/01023
; PRIOR FILING DATE: 1998-12-10
; PRIOR APPLICATION NUMBER: AU PP 0839
; PRIOR FILING DATE: 1997-12-10
; PRIOR APPLICATION NUMBER: AU PP 1182
; PRIOR FILING DATE: 1997-12-31
; PRIOR APPLICATION NUMBER: AU PP 1846
; PRIOR FILING DATE: 1998-01-30
; PRIOR APPLICATION NUMBER: AU PP 2264
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: AU PP 2911
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: AU PP 3128
; PRIOR FILING DATE: 1998-04-23
; PRIOR APPLICATION NUMBER: AU PP 3338
; PRIOR FILING DATE: 1998-05-05
; PRIOR APPLICATION NUMBER: AU PP 3654
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: AU PP 4917
; PRIOR FILING DATE: 1998-07-29
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 721
; SOFTWARE: Patent version 3.2
; SEQ ID NO 424
; LENGTH: 1017
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-09-581-286A-424

Query Match 100.0%; Score 1017; DB 5; Length 1017;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1017; Conservative 0; Mismatches 0; Gaps 0;

Qy 1 MKRMTLFFCLLTISIGWMAQNRVTGVTIISSEDNEPLIGANVVVGVNTTIGATDLGDN 60
 Db 1 MKRMTLFFCLLTISIGWMAQNRVTGVTIISSEDNEPLIGANVVVGVNTTIGATDLGDN 60
 Qy 61 FTLSVPANAKMLRVSYSGMTTKEVAIANVWKIVLDPDSKVLQVVLGYGTGQKLSVSG 120
 Db 61 FTLSVPANAKMLRVSYSGMTTKEVAIANVWKIVLDPDSKVLQVVLGYGTGQKLSVSG 120
 Qy 121 SVAKVSSEKLAEPVANIIMDALQOVAGQVMTTSGDPTAVASVEIHGTGSLGASSAPLY 180
 Db 121 SVAKVSSEKLAEPVANIIMDALQOVAGQVMTTSGDPTAVASVEIHGTGSLGASSAPLY 180
 Qy 181 IVDMQTSLDVATWMPNDFESMSVLKASATSIIYGARAANGVVFIOTKKGKMSERGRIT 240
 Db 181 IVDMQTSLDVATWMPNDFESMSVLKASATSIIYGARAANGVVFIOTKKGKMSERGRIT 240
 Qy 241 FNASYGISQILNTKPLDNMTGDBELLDFQVKGAGFWGNNTVQVKWMLAGAEADLYGNYD 300
 Db 241 FNASYGISQILNTKPLDNMTGDBELLDFQVKGAGFWGNNTVQVKWMLAGAEADLYGNYD 300
 Qy 301 SLKDEYKTLFPVDFNHDAWMLKALFKTAPTSQDISFSGSGSQTYSYASIGYFDQEGMA 360
 Db 301 SLKDEYKTLFPVDFNHDAWMLKALFKTAPTSQDISFSGSGSQTYSYASIGYFDQEGMA 360
 Qy 361 REPANFKRYSGRLNPFESRINEMWKVGANLSGAIANRRSADYFGKYNGSGTFFGLTMPRY 420
 Db 361 REPANFKRYSGRLNPFESRINEMWKVGANLSGAIANRRSADYFGKYNGSGTFFGLTMPRY 420
 Qy 421 YNPDPVNGDLADVYMTGATRPSTPEYPAKMRPFSSSHQANVNGPAQITPIKGLTLKA 480
 Db 421 YNPDPVNGDLADVYMTGATRPSTPEYPAKMRPFSSSHQANVNGPAQITPIKGLTLKA 480
 Qy 481 QAGVDIINTTSSKRMNNPVDSTPLGERERAYRDVSKFTNTAEYKFSIDEXHDLTAL 540
 Db 481 QAGVDIINTTSSKRMNNPVDSTPLGERERAYRDVSKFTNTAEYKFSIDEXHDLTAL 540
 Qy 541 MGHEYIEYEGDVGASGKSPESDKMLLSQKTKGNSLSLPEHRVAEYAYLSFFSRFNYGF 600
 Db 541 MGHEYIEYEGDVGASGKSPESDKMLLSQKTKGNSLSLPEHRVAEYAYLSFFSRFNYGF 600
 Qy 601 DKWYIDFSVRNDQSSRFGSNRRSAWFSYVGGMFDIYNKFIQESNWLSDLRLKMSYGTG 660
 Db 601 DKWYIDFSVRNDQSSRFGSNRRSAWFSYVGGMFDIYNKFIQESNWLSDLRLKMSYGTG 660
 Qy 661 NSEIGNYHQALVTNNYTEDAMGLSISTAGNPDLSWEKOSQFNFLAAGAFNNRLSAEV 720
 Db 661 NSEIGNYHQALVTNNYTEDAMGLSISTAGNPDLSWEKOSQFNFLAAGAFNNRLSAEV 720
 Qy 721 DFYVRTTNDMLIDVPMPIYISGFFSQYQNVGSMKNTGVDLSLKGTYIQNKDWNVYASAFN 780
 Db 721 DFYVRTTNDMLIDVPMPIYISGFFSQYQNVGSMKNTGVDLSLKGTYIQNKDWNVYASAFN 780
 Qy 781 YNRQETIKLFFGLNKYMLPNTGTIWEIGYPNSFTMAEYAGIDKKTGKQLWYVPGQVDAG 840
 Db 781 YNRQETIKLFFGLNKYMLPNTGTIWEIGYPNSFTMAEYAGIDKKTGKQLWYVPGQVDAG 840
 Qy 841 NKVTTQSYADLETRIDKSVTPPTTGGFSIGASKWGLSLDADFAIYVKGKMINNDRYFTE 900
 Db 841 NKVTTQSYADLETRIDKSVTPPTTGGFSIGASKWGLSLDADFAIYVKGKMINNDRYFTE 900
 Qy 901 NAGGLMQLNDKMLNNAWTEENKEDTDPKLGQSFPQFOTHLLENASFLRLKMLKLYVLPN 960
 Db 901 NAGGLMQLNDKMLNNAWTEENKEDTDPKLGQSFPQFOTHLLENASFLRLKMLKLYVLPN 960
 Qy 961 SLFAGQNVIGARVYLMARNLLTVTKYKGFDPPEAGNVGNQYPSNQYVAGIQLSF 1017
 Db 961 SLFAGQNVIGARVYLMARNLLTVTKYKGFDPPEAGNVGNQYPSNQYVAGIQLSF 1017

RESULT 2

US-09-581-286A-300

; Sequence 300, Application US/09581286A

; GENERAL INFORMATION:

; APPLICANT: ROSS, BRUCE C.
 ; APPLICANT: BARR, IAN G.
 ; APPLICANT: PATTERSON, MICHELLE A.
 ; APPLICANT: AGIUS, CATHERINE T.
 ; APPLICANT: ROTHSEL, LINDA J.
 ; APPLICANT: MARGETTS, MAL B.
 ; APPLICANT: HOCKING, DIANNA M.
 ; APPLICANT: WEBB, ELIZABETH A.
 ; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS POLYPEPTIDES AND NUCLEOTIDES
 ; FILE REFERENCE: 4137-3
 ; CURRENT APPLICATION NUMBER: US/09/581,286A
 ; CURRENT FILING DATE: 2000-06-28
 ; PRIOR APPLICATION NUMBER: PCT/AU98/01023
 ; PRIOR FILING DATE: 1998-12-10
 ; PRIOR APPLICATION NUMBER: AU PP 0839
 ; PRIOR FILING DATE: 1997-12-10
 ; PRIOR APPLICATION NUMBER: AU PP 1182
 ; PRIOR FILING DATE: 1997-12-31
 ; PRIOR APPLICATION NUMBER: AU PP 1846
 ; PRIOR FILING DATE: 1998-01-30
 ; PRIOR APPLICATION NUMBER: AU PP 2264
 ; PRIOR FILING DATE: 1998-03-10
 ; PRIOR APPLICATION NUMBER: AU PP 2911
 ; PRIOR FILING DATE: 1998-04-09
 ; PRIOR APPLICATION NUMBER: AU PP 3128
 ; PRIOR FILING DATE: 1998-04-23
 ; PRIOR APPLICATION NUMBER: AU PP 3338
 ; PRIOR FILING DATE: 1998-05-05
 ; PRIOR APPLICATION NUMBER: AU PP 3654
 ; PRIOR FILING DATE: 1998-05-22
 ; PRIOR APPLICATION NUMBER: AU PP 4917
 ; PRIOR FILING DATE: 1998-07-29
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 721
 ; SOFTWARE: Patent in version 3.2
 ; SEQ ID NO 300
 ; LENGTH: 1046
 ; TYPE: PRT
 ; ORGANISM: Porphyromonas gingivalis
 ; US-09-581-286A-300

Query Match 100.0%; Score 1017; DB 5; Length 1046;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1017; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKRMTLFFCLLTISIGWMAQNRVTGVTIISSEDNEPLIGANVVVGVNTTIGATDLGDN 60
 Db 30 MKRMTLFFCLLTISIGWMAQNRVTGVTIISSEDNEPLIGANVVVGVNTTIGATDLGDN 89
 Qy 61 FTLSVPANAKMLRVSYSGMTTKEVAIANVWKIVLDPDSKVLQVVLGYGTGQKLSVSG 120
 Db 90 FTLSVPANAKMLRVSYSGMTTKEVAIANVWKIVLDPDSKVLQVVLGYGTGQKLSVSG 149
 Qy 121 SVAKVSSEKLAEPVANIIMDALQOVAGQVMTTSGDPTAVASVEIHGTGSLGASSAPLY 180
 Db 150 SVAKVSSEKLAEPVANIIMDALQOVAGQVMTTSGDPTAVASVEIHGTGSLGASSAPLY 209
 Qy 181 IVDMQTSLDVATWMPNDFESMSVLKASATSIIYGARAANGVVFIOTKKGKMSERGRIT 240
 Db 210 IVDMQTSLDVATWMPNDFESMSVLKASATSIIYGARAANGVVFIOTKKGKMSERGRIT 269
 Qy 241 FNASYGISQILNTKPLDNMTGDBELLDFQVKGAGFWGNNTVQVKWMLAGAEADLYGNYD 300
 Db 270 FNASYGISQILNTKPLDNMTGDBELLDFQVKGAGFWGNNTVQVKWMLAGAEADLYGNYD 329
 Qy 301 SLKDEYKTLFPVDFNHDAWMLKALFKTAPTSQDISFSGSGSQTYSYASIGYFDQEGMA 360
 Db 330 SLKDEYKTLFPVDFNHDAWMLKALFKTAPTSQDISFSGSGSQTYSYASIGYFDQEGMA 389
 Qy 361 REPANFKRYSGRLNPFESRINEMWKVGANLSGAIANRRSADYFGKYNGSGTFFGLTMPRY 420
 Db 390 REPANFKRYSGRLNPFESRINEMWKVGANLSGAIANRRSADYFGKYNGSGTFFGLTMPRY 449

QY 421 YNPPDVNGDLADVYMYGATRSMTPEYPAKMRPFSSSHQANVNGFAQITPIKGLTLKA 480
Db 450 YNPPDVNGDLADVYMYGATRSMTPEYPAKMRPFSSSHQANVNGFAQITPIKGLTLKA 509
QY 481 QAGVDITNTRTSSKMPNPNYDSTPLGERERAYRDVSKSFNTAEYKFSIDEXHDLTAL 540
Db 510 QAGVDITNTRTSSKMPNPNYDSTPLGERERAYRDVSKSFNTAEYKFSIDEXHDLTAL 569
QY 541 MGHEIYEGDVIGASSKGFESDKMLLSQGTGNSLSLPEHRVAEYAYLSFFSRFNYGF 600
Db 570 MGHEIYEGDVIGASSKGFESDKMLLSQGTGNSLSLPEHRVAEYAYLSFFSRFNYGF 629
QY 601 DKWYIDFSVRNDQSRFGSNRRSAWFSYVGMFDIYKFTIOESNWLSDLRLKMSYGTG 660
Db 630 DKWYIDFSVRNDQSRFGSNRRSAWFSYVGMFDIYKFTIOESNWLSDLRLKMSYGTG 689
QY 661 NSEIGNYHQAALVTNNYTEDAMGLSISTAGNPDLSEKQSFNFGLAAGAFNNLSAEV 720
Db 690 NSEIGNYHQAALVTNNYTEDAMGLSISTAGNPDLSEKQSFNFGLAAGAFNNLSAEV 749
QY 721 DFYVRTNDMLLDVPMPIYSGFSSQYQNVGSMKNTGVDLSLKGTYQNKDMNVYASAFN 780
Db 750 DFYVRTNDMLLDVPMPIYSGFSSQYQNVGSMKNTGVDLSLKGTYQNKDMNVYASAFN 809
QY 781 YNRQBITKLFGLNKYMLPNTGTIWEIGYPNSFYMAEYAGIDKKTGKQLWYVPGQVDAG 840
Db 810 YNRQBITKLFGLNKYMLPNTGTIWEIGYPNSFYMAEYAGIDKKTGKQLWYVPGQVDAG 869
QY 841 NKVTTSOYADLETRIDKSVTPITGGPSLGASWGLSLDDPAVYVGVKMMINNDRYTE 900
Db 870 NKVTTSOYADLETRIDKSVTPITGGPSLGASWGLSLDDPAVYVGVKMMINNDRYTE 929
QY 901 NAGGLMQLMKMLNNAWTEDEKTDVPLKQSPQFDTHLENASFLRLKMLTYVLPN 960
Db 930 NAGGLMQLMKMLNNAWTEDEKTDVPLKQSPQFDTHLENASFLRLKMLTYVLPN 989
QY 961 SLFAGQNVIGARVYLMARNLLTVTKYKGFDPGAGNVGKNQYPSKQYVAGIQLSF 1017
Db 990 SLFAGQNVIGARVYLMARNLLTVTKYKGFDPGAGNVGKNQYPSKQYVAGIQLSF 1046

RESULT 3
US-09-581-286A-425
; Sequence 425, Application US/09581286A
; GENERAL INFORMATION:
; APPLICANT: ROSS, BRUCE C.
; APPLICANT: BARR, TAN G.
; APPLICANT: PATTERSON, MICHELLE A.
; APPLICANT: AGIUS, CATHERINE T.
; APPLICANT: ROTH, LINDA J.
; APPLICANT: MARGRETT, MAL B.
; APPLICANT: HOCKING, DIANNA M.
; APPLICANT: WEBB, ELIZABETH A.
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS POLYPEPTIDES AND NUCLEOTIDES
; FILE REFERENCE: 4137-3
; CURRENT APPLICATION NUMBER: US/09/581,286A
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: PCT/AU98/01023
; PRIOR FILING DATE: 1998-12-10
; PRIOR APPLICATION NUMBER: AU PP 0839
; PRIOR FILING DATE: 1997-12-10
; PRIOR APPLICATION NUMBER: AU PP 1182
; PRIOR FILING DATE: 1997-12-31
; PRIOR APPLICATION NUMBER: AU PP 1846
; PRIOR FILING DATE: 1998-01-30
; PRIOR APPLICATION NUMBER: AU PP 2264
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: AU PP 2911
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: AU PP 3128
; PRIOR FILING DATE: 1998-04-23
; PRIOR APPLICATION NUMBER: AU PP 3338
; PRIOR FILING DATE: 1998-05-05

; PRIOR APPLICATION NUMBER: AU PP 3654
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: AU PP 4917
; PRIOR FILING DATE: 1998-07-29
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 721
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 425
; LENGTH: 1014
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
; US-09-581-286A-425

Query Match 99.7%; Score 1014; DB 5; Length 1014;
Best Local Similarity 100.0%; Pred.No. 0;
Matches 1014; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 MTLFPLCLLTISIGWAMAQNRTVKTIVISSENEPLIGANVVVGVNTTGAATDLGNFTL 63
Db 1 MTLFPLCLLTISIGWAMAQNRTVKTIVISSENEPLIGANVVVGVNTTGAATDLGNFTL 60
QY 64 SVPANAKMLRYSYSGMTTKEVAIANVMKIVLDDPSKVLQVVLGVGTGOKLSTVSGSVA 123
Db 61 SVPANAKMLRYSYSGMTTKEVAIANVMKIVLDDPSKVLQVVLGVGTGOKLSTVSGSVA 120
QY 124 KVSSEKLAKEPVAANIMDALQOVAGMQVMTTSGDPTAVASVEIHGTGSLGASSAPLYIVD 183
Db 121 KVSSEKLAKEPVAANIMDALQOVAGMQVMTTSGDPTAVASVEIHGTGSLGASSAPLYIVD 180
QY 184 QMOTSLDVATWNPNDPESMSVLKDAATSISYGAARAANGVVFPIQTKKMSERGRITNA 243
Db 181 QMOTSLDVATWNPNDPESMSVLKDAATSISYGAARAANGVVFPIQTKKMSERGRITNA 240
QY 244 SYGISOILNTPLDNMMTGDELLDFQVKAGFWGNNQTVQVKQKMDTLAGAEDLYGNYDSLK 303
Db 241 SYGISOILNTPLDNMMTGDELLDFQVKAGFWGNNQTVQVKQKMDTLAGAEDLYGNYDSLK 300
QY 304 DEYKGTLPFVDFNHDADWLKALFKTAPTSQGDISFSGSGSQTSYVASYGYDQEGMAREP 363
Db 301 DEYKGTLPFVDFNHDADWLKALFKTAPTSQGDISFSGSGSQTSYVASYGYDQEGMAREP 360
QY 364 ANFKRYSGRLNPFESRINEWLKVGANLSGAIANRRSADYFGKYMGSGTGGVLTMPRYNYP 423
Db 361 ANFKRYSGRLNPFESRINEWLKVGANLSGAIANRRSADYFGKYMGSGTGGVLTMPRYNYP 420
QY 424 FVNGDLADVYMYGATRSMTPEYPAKMRPFSSSHQANVNGFAQITPIKGLTLKAQAG 483
Db 421 FVNGDLADVYMYGATRSMTPEYPAKMRPFSSSHQANVNGFAQITPIKGLTLKAQAG 480
QY 484 VDI TNTRTSSKMPNPNYDSTPLGERERAYRDVSKSFNTAEYKFSIDEXHDLTALMGH 543
Db 481 VDI TNTRTSSKMPNPNYDSTPLGERERAYRDVSKSFNTAEYKFSIDEXHDLTALMGH 540
QY 544 EYIEYEGDVIGASSKGFESDKMLLSQGTGNSLSLPEHRVAEYAYLSFFSRFNYGFDPKW 603
Db 541 EYIEYEGDVIGASSKGFESDKMLLSQGTGNSLSLPEHRVAEYAYLSFFSRFNYGFDPKW 600
QY 604 MYIDFSVRNDQSRFGSNRRSAWFSYVGMFDIYKFTIOESNWLSDLRLKMSYGTGNS 663
Db 601 MYIDFSVRNDQSRFGSNRRSAWFSYVGMFDIYKFTIOESNWLSDLRLKMSYGTGNS 660
QY 664 IGNYHQAALVTNNYTEDAMGLSISTAGNPDLSEKQSFNFGLAAGAFNNLSAEVDPY 723
Db 661 IGNYHQAALVTNNYTEDAMGLSISTAGNPDLSEKQSFNFGLAAGAFNNLSAEVDPY 720
QY 724 VRTNDMLLDVPMPIYSGFSSQYQNVGSMKNTGVDLSLKGTYQNKDMNVYASAFN 783
Db 721 VRTNDMLLDVPMPIYSGFSSQYQNVGSMKNTGVDLSLKGTYQNKDMNVYASAFN 780
QY 784 QEITKLPFLNKYMLPNTGTIWEIGYPNSFYMAEYAGIDKKTGKQLWYVPGQVDAGNKV 843
Db 781 QEITKLPFLNKYMLPNTGTIWEIGYPNSFYMAEYAGIDKKTGKQLWYVPGQVDAGNKV 840

Qy 844 TTQSYADLETRIDKSVTPITGGTSGAGSKGLSLDADPAYIVGKWMNNDRYFTENAG 903
 Db 841 TTQSYADLETRIDKSVTPITGGTSGAGSKGLSLDADPAYIVGKWMNNDRYFTENAG 900
 Qy 904 GLMQLNKDKMLLNAWTENDKEDTVPKLGQSPQFPTHLENASFLRLKLNKLTYYVLPNSLF 963
 Db 901 GLMQLNKDKMLLNAWTENDKEDTVPKLGQSPQFPTHLENASFLRLKLNKLTYYVLPNSLF 960
 Qy 964 AGQNVIGGARVYLMARNLITVTKYKGFDPGAGNVGNKQYVAGIQLSF 1017
 Db 961 AGQNVIGGARVYLMARNLITVTKYKGFDPGAGNVGNKQYVAGIQLSF 1014

RESULT 4
 US-09-597-516A-5632
 ; Sequence 5632, Application US/09897516A
 ; GENERAL INFORMATION:
 ; APPLICANT: Corbin, David R.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Huesing, Joseph E.
 ; APPLICANT: Malvar, Thomas M.
 ; APPLICANT: Krasomil-Osterfeld, Karina C.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Spiridonov, Sergei
 ; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
 ; FILE REFERENCE: 38-21(51847)B
 ; CURRENT APPLICATION NUMBER: US/09/897,516A
 ; CURRENT FILING DATE: 2001-06-29
 ; PRIOR APPLICATION NUMBER: US 60/215,161
 ; PRIOR FILING DATE: 2000-06-30
 ; NUMBER OF SEQ ID NOS: 8415
 ; SEQ ID NO 5632
 ; LENGTH: 389
 ; TYPE: PRT
 ; ORGANISM: Xenorhabdus sp.
 US-09-597-516A-5632

Query Match 0.8%; Score 8; DB 5; Length 389;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 127 SEKLAEXP 134
 Db 270 SEKLAEXP 277

RESULT 5
 US-10-472-928-880
 ; Sequence 880, Application US/10472928
 ; GENERAL INFORMATION:
 ; APPLICANT: CHIRON spa
 ; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
 ; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
 ; FILE REFERENCE: P026926W0
 ; CURRENT APPLICATION NUMBER: US/10/472,928
 ; CURRENT FILING DATE: 2003-09-26
 ; PRIOR APPLICATION NUMBER: GB-0107658.7
 ; PRIOR FILING DATE: 2001-03-27
 ; NUMBER OF SEQ ID NOS: 4979
 ; SOFTWARE: SeqWin99, version 1.03
 ; SEQ ID NO 880
 ; LENGTH: 487
 ; TYPE: PRT
 ; ORGANISM: Streptococcus pneumoniae
 ; FEATURE:
 ; OTHER INFORMATION: type I restriction-modification system, M subunit (hsdM)
 ; OTHER INFORMATION: Cellular location: cytoplasm
 ; OTHER INFORMATION: Similar to strain R6 sequence 15902493 (0.E+01)
 US-10-472-928-880

Query Match 0.8%; Score 8; DB 6; Length 487;
 Best Local Similarity 100.0%; Pred. No. 19;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 261 TGDLELDF 268
 Db 77 TGDLELDF 84

RESULT 6
 US-10-679-063-14822
 ; Sequence 14822, Application US/10679063
 ; GENERAL INFORMATION:
 ; APPLICANT: Edgerton, Michael D
 ; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
 ; FILE REFERENCE: 38-15(52054)B
 ; CURRENT APPLICATION NUMBER: US/10/679,063
 ; CURRENT FILING DATE: 2003-10-02
 ; PRIOR APPLICATION NUMBER: 60/415,758
 ; PRIOR FILING DATE: 2002-10-02
 ; NUMBER OF SEQ ID NOS: 27373
 ; SEQ ID NO 14822
 ; LENGTH: 1127
 ; TYPE: PRT
 ; ORGANISM: Xanthomonas axonopodis pv. citri str. 306
 US-10-679-063-14822

Query Match 0.8%; Score 8; DB 6; Length 1127;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 170 GSLGASSA 177
 Db 949 GSLGASSA 956

RESULT 7
 PCT-US03-05499-28
 ; Sequence 28, Application PC/TUS0305499
 ; GENERAL INFORMATION:
 ; APPLICANT: APOVIA, INC.
 ; APPLICANT: Peck, Birgit
 ; APPLICANT: Ashley, Birkett J.
 ; TITLE OF INVENTION: STABILIZED HBC CHIMER PARTICLES HAVING MENINGOCOCCAL IMMUNOGENS
 ; FILE REFERENCE: ICC-107 PCT (4564/88525)
 ; CURRENT APPLICATION NUMBER: PCT/US03/05499
 ; CURRENT FILING DATE: 2003-02-21
 ; PRIOR APPLICATION NUMBER: US 60/358,618
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 185
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 28
 ; LENGTH: 18
 ; TYPE: PRT
 ; ORGANISM: Neisseria meningitidis
 PCT-US03-05499-28

Query Match 0.7%; Score 7; DB 1; Length 18;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 569 SQCKTGN 575
 Db 5 SQCKTGN 11

RESULT 8
 US-10-677-074-105
 ; Sequence 105, Application US/10677074
 ; GENERAL INFORMATION:
 ; APPLICANT: Page, Mark
 ; APPLICANT: Friede, Martin
 ; APPLICANT: Schmidt, Annette Elisabeth
 ; APPLICANT: Stober, Detlef
 ; TITLE OF INVENTION: STABILIZED HBC CHIMER PARTICLES AS THERAPEUTIC VACCINE FOR


```
; TITLE OF INVENTION: CHRONIC HEPATITIS
; FILE REFERENCE: 4564/87179
; CURRENT APPLICATION NUMBER: US/10/677,074
; CURRENT FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: 10/372,076
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 10/080,299
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 10/082,014
; PRIOR FILING DATE: 2002-02-22
; NUMBER OF SEQ ID NOS: 308
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 105
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-10-677-074-105

Query Match      0.7%; Score 7; DB 6; Length 18;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 569 SQKGTG 575
Db 5 SQKGTG 11

RESULT 9
US-60-519-270-199
; Sequence 199, Application US/60519270
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: BEGOVICH, Ann
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01490
; CURRENT APPLICATION NUMBER: US/60/519,270
; CURRENT FILING DATE: 2003-11-13
; NUMBER OF SEQ ID NOS: 14048
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 199
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-519-270-199

Query Match      0.7%; Score 7; DB 7; Length 37;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 785 EITKLF 791
Db 6 EITKLF 12

RESULT 10
US-10-425-114A-48932
; Sequence 48932, Application US/10425114A
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114A
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 48932
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Zea mays
```

```
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3354-020-B12_FLI.pgp
US-10-425-114A-48932

Query Match      0.7%; Score 7; DB 6; Length 117;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 130 LAKEPVA 136
Db 48 LAKEPVA 54

RESULT 11
US-10-679-063-18798
; Sequence 18798, Application US/10679063
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52054)B
; CURRENT APPLICATION NUMBER: US/10/679,063
; CURRENT FILING DATE: 2003-10-02
; PRIOR APPLICATION NUMBER: 60/415,758
; PRIOR FILING DATE: 2002-10-02
; NUMBER OF SEQ ID NOS: 27373
; SEQ ID NO 18798
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Streptococcus agalactiae NEM316
US-10-679-063-18798

Query Match      0.7%; Score 7; DB 6; Length 127;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 890 WWINNDR 896
Db 21 WWINNDR 27

RESULT 12
US-10-425-114A-45140
; Sequence 45140, Application US/10425114A
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114A
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 45140
; LENGTH: 134
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700166824_FLI.pgp
US-10-425-114A-45140

Query Match      0.7%; Score 7; DB 6; Length 134;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 894 NDRYFTE 900
Db 69 NDRYFTE 75
```

RESULT 13

US-10-425-114A-69574

; Sequence 69574, Application US/10425114A
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114A
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 69574
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73072D10_FLI.pep
US-10-425-114A-69574

Query Match 0.7%; Score 7; DB 6; Length 138;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 864 ITGGFSL 870
|||||
DB 68 ITGGFSL 74

RESULT 14

US-10-425-114A-57907

; Sequence 57907, Application US/10425114A
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114A
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 57907
; LENGTH: 141
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17014D12_FLI.pep
US-10-425-114A-57907

Query Match 0.7%; Score 7; DB 6; Length 141;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 752 MKNTGVD 758
|||||
DB 35 MKNTGVD 41

RESULT 15

US-10-425-114A-67079

; Sequence 67079, Application US/10425114A
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong